LIBRARY

CURRENT YR/VOL
arygrove College Library
425 West McNichols Road
Detroit, MI 48221

FEBRUARY 2010 • VOLUME 81, NUMBER 1

CHILD DEVELOPMENT

Editor
JEFFREY J. LOCKMAN

Associate Editors
MARGARET BURCHINAL • JOHN COLOMBO • MICHAEL CUNNINGHAM • MARY DOZIER
ANDREW J. FULIGNI • MARY GAUVAIN • ROBERTA M. GOLINKOFF • NANCY G. GUERRA
D. GEOFFREY HALL • BETSY HOZA • MELANIE KILLEN • LOUIS J. MOSES • CYNTHIA STIFTER

SPECIAL SECTION
The Effects of Early Experience on Development

Special Section Editors
NATHAN A. FOX • MICHAEL RUTTER

Society for Research
in Child Development

Published by Wiley-Blackwell for the Society for Research in Child Development

Published by Wiley-Blackwell for the Society for Research in Child Development

Child Development

CHILD DEVELOPMENT (ISSN 0009-3920), one of three publications of the Society for Research in Child Development, Inc., is published bimonthly by Wiley Subscription Services, Inc., a Wiley Company, 111 River St., Hoboken, NJ 07030-5774. Periodical Postage Paid at Hoboken, NJ and additional offices.

Postmaster: Send all address changes to *CHILD DEVELOPMENT*, Journal Customer Services, John Wiley & Sons Inc., 350 Main Street, Malden, MA 02148-5020.

Information for Subscribers: *Child Development* is published in six issues per year. Institutional subscription prices for 2010 are: Print & Online: US$670 (US), US$938 (Rest of World), €608 (Europe), £479 (UK). Prices are exclusive of tax. Asia- Pacific GST, Canadian GST and European VAT will be applied at the appropriate rates. For more information on current tax rates, please go to www3.interscience.wiley.com/aboutus/journal_ordering_and_payment.html#Tax. The price includes online access to the current and all online back files to January 1st 1997, where available. For other pricing options, including access information and terms and conditions, please visit www.interscience.wiley.com/ journal-info.

Delivery Terms and Legal Title: Prices include delivery of print journals to the recipient's address. Delivery terms are Delivered Duty Unpaid (DDU); the recipient is responsible for paying any import duty or taxes. Legal title passes to the customer on despatch by our distributors.

For more information about Wiley-Blackwell journals, including online access information, terms and conditions, and other pricing options, please visit www.interscience.wiley.com or contact our customer service department, tel: (800) 835-6770 or +1 (781) 388-8598 (Americas); +44 (0) 1865 778315 (Europe, Middle East, and Africa); or +65 6511 8000 (Asia Pacific). Subscription in 2010 includes three issues of *Monographs of the Society for Research in Child Development* and three issues of *Child Development Perspectives*. Single issues from current and recent volumes are available at the current single issue price from csjournals@wiley.com. Earlier issues may be obtained from Periodicals Service Company, 11 Main Street, Germantown, NY 12526, USA. Tel: +1 518 537 4700, Fax: +1 518 537 5899, Email: psc@periodicals.com. Volumes 37 to date are available in microfilm from ProQuest Information and Learning, Serials Customer Service Department, 300 North Zeeb Road, Ann Arbor, MI 48106; telephone: (800) 521-0600 x2873.

Electronic Access: The *Child Development* website can be found at www.srcd.org. Abstract information for this journal is electronically available at www.blackwellpublishing.com/cdev. This journal is available online at Wiley InterScience. Visit www3.interscience.wiley.com to search the articles and register for table of contents e-mail alerts.

Access to this journal is available free online within institutions in the developing world through the HINARI initiative with the WHO. For information, visit www.healthinternetwork.org

Journal Customer Services: For ordering information, claims and any enquiry concerning your journal subscription, please go to interscience.wiley.com/support or contact your nearest office:

Americas: Email: cs-journals@wiley.com; Tel: +1 781 388 8598 or 1 800 835 6770 (Toll free in the USA & Canada).
Europe, Middle East, and Africa: Email: cs-journals@wiley.com; Tel: +44 (0) 1865 778315
Asia Pacific: Email: cs-journals@wiley.com; Tel: +65 6511 8000
Japan: For Japanese speaking support: Email: cs-japan@wiley.com; Tel: +65 6511 8010 or Tel (toll-free): 005 316 50 480.
Further Japanese customer support is also available at www.interscience.wiley.com/support
Visit our Online Customer Self-Help available in 6 languages at www.interscience.wiley.com/support

Membership communication and address changes should be addressed to the SRCD Executive Office, 2950 South State Street, Suite 401, Ann Arbor, MI 48104, USA. http://www.srcd.org/

Copyright and Photocopying: Journal Compilation © 2010 The Society for Research in Child Development, Inc. All rights reserved. No part of this publication (apart from articles marked "OnlineOpen") may be reproduced, stored or transmitted in any form or by any means without the prior permission in writing from the copyright holder. Authorization to photocopy items for internal and personal use is granted by the copyright holder for libraries and other users registered with their local Reproduction Rights Organization (RRO), e.g., Copyright Clearance Center (CCC), 222 Rosewood Drive, Danvers, MA 01923, USA (www.copyright.com), provided the appropriate fee is paid directly to the RRO. This consent does not extend to other kinds of copying such as copying for general distribution, for advertising and promotional purposes, for creating new collective works or for resale. OnlineOpen articles are made available in accordance with the terms of the Creative Commons Deed, Attribution 2.5 (further details available from www.creativecommons.org), which allows Open Access dissemination of the article, but does not permit commercial exploitation or the creation of derivative works. Special requests should be addressed to PermissionsUK@wiley.com.

Production Editor: Teresa Huang (email: CDEV@wiley.com).

Advertising: Kristin McCarthy (email: kmccarthy@wiley.com).

Indexing/Abstracting: The contents of this journal are indexed or abstracted in the following: *Academic ASAP; Expanded Academic ASAP; Academic Search Elite; Academic Search Premier; Applied Social Sciences Index and Abstracts; Biomedical Reference Collection: Comprehensive Edition; CAB Abstracts; CAB Direct; Contents Pages in Education; Current Contents/Social & Behavioral Science; Current Index to Journals in Education; Cambridge Scientific Abstracts Linguistics & Language Behavior Abstracts; CatchWord; Current Index to Journals in Education; Educational Research Abstracts; Environmental Sciences & Pollution Management; e-psyche; EBSCO Sociological Collection; EBSCO Online; EBSCO Professional Development Collection; Educational Research Abstracts; Educational Resources Information Center; Health and Safety Science Abstracts; ISI Basic Social Sciences Index; InfoTrac OneFile; Ingenta; JCR Social Sciences Edition; Leisure, Recreation, and Tourism Abstracts; Online Computer Library Center FirstSearch Electronic Collections Online; Periodical Abstracts Library; Periodical Abstracts Research II; ProQuest 5000; ProQuest Social Science Plus Text; PsycINFO; Psychological Abstracts; Psychology and Behavioral Sciences Collection; Pollution Abstracts; Safety Science & Risk Abstracts; Sage Family Studies Abstracts; Social Services Abstracts; Social Science Citation Index; Sociological Abstracts; Student Resource Center Bronze; Student Resource Center Gold; Student Resource Center Silver; Violence & Abuse Abstracts; Wilson OmniFile V; Wilson Social Sciences, Abstracts; and Wilson Social Sciences Index*

Published by Wiley-Blackwell for the Society for Research in Child Development

Child Development

Editor

Jeffrey J. Lockman
Tulane University

Associate Editors

Margaret Burchinal, *University of North Carolina at Chapel Hill*
John Colombo, *University of Kansas*
Michael Cunningham, *Tulane University*
Mary Dozier, *University of Delaware*
Andrew J. Fuligni, *University of California, Los Angeles*
Mary Gauvain, *University of California, Riverside*
Roberta M. Golinkoff, *University of Delaware*
Nancy G. Guerra, *University of California, Riverside*
D. Geoffrey Hall, *University of British Columbia*
Betsy Hoza, *University of Vermont*
Melanie Killen, *University of Maryland*
Louis J. Moses, *University of Oregon*
Cynthia Stifter, *The Pennsylvania State University*

Managing Editor

Detra N. Davis

Editorial Assistant

Antonella Caiazza

Consulting Editors

Emma Adam
Karen E. Adolph
Nazan Aksan
William F. Arsenio
Daniel H. Ashmead
Anna Axia
Karen Bartsch
Annie Bernier
Susan Birch
Amy Booth
Heather Bouchey
Paul Boxer
Gregory Stephen Braswell
Mara Brendgen
Gene H. Brody
Keith Burt
Emily W. Bushnell
Gustavo Carlo
Stephanie M. Carlson
Ruth K. Chao
Enrique Chaux
Pablo Chavajay
Tabbye Maria Chavous
Elysia Davis
David Dickinson
Gil Diesendruck
Ruth Feldman
Lisa Gatzke-Kopp
Tamsin C. German
Rick O. Gilmore
Artin Goncu
Susan A. Graham
Elena L. Grigorenko
Laura D. Hanish
Daniel Hart
David B. Henry
Claire H. Hughes
Mutsumi Imai
Deborah J. Johnson
Roger Kobak
Joachim Krueger
Richard Lee
Charlie Lewis
Jeffrey Lidz
Tina Malti
Jelani Mandara
Gary Marcus
Ann S. Masten
Bob McMurray
Julie Meeks-Gardner
Jayanthi Mistry
Ginger A. Moore
Dianna (Annie) Murray-Close
Velma McBride Murry
Letitia Naigles
Drew Nesdale
Nora S. Newcombe
David Oppenheim
Ana Papafragou
Susan M. Perez
Bradford Pillow
Eva Pomerantz
Douglas Powell
Kimberly K. Powlishta
Samuel Putnam
Suzanne M. Randolph
Robert Reeve
Kristen D. Ritchey
John J. Rieser
James L. Rodriguez
Glenn I. Roisman
Susan A. Rose
Martin D. Ruck
Jenny Saffran
Carlo Schuengel
David Schwartz
Judi G. Smetana
Alan L. Smith
Susan J. Spieker
Timothy Stickle
Dena Phillips Swanson
Adriana Umaña-Taylor
Christopher Trentacosta
Kimberly Updegraff
Marinus H. van IJzendoorn
Brian E. Vaughn
Elisa Velasquez
Lynne Vernon-Feagans
Peter Vishton
Elizabeth Votruba-Drzal
Cecilia Wainryb
Sarah Watamura
Henry M. Wellman
Mike Willoughby
Philip David Zelazo

Child Development, January/February 2010, Volume 81, Number 1, Pages 1–429

Contents

IN THIS ISSUE 1

SRCD PRESIDENTIAL ADDRESS

A Unified Theory of Development: A Dialectic Integration of Nature and Nurture
Arnold Sameroff 6

SPECIAL SECTION

The Effects of Early Experience on Development
Editors: Nathan A. Fox and Michael Rutter

INTRODUCTION

Introduction to the Special Section on The Effects of Early Experience on Development
Nathan A. Fox and Michael Rutter 23

SPECIAL SECTION REVIEWS

How the Timing and Quality of Early Experiences Influence the Development of Brain Architecture
Sharon E. Fox, Pat Levitt, and Charles A. Nelson III 28

Epigenetics and the Biological Definition of Gene × Environment Interactions
Michael J. Meaney 41

SPECIAL SECTION EMPIRICAL ARTICLES

Fetal Development and Early Experience

A Quasi-Experimental Study of Maternal Smoking During Pregnancy and Offspring Academic Achievement
Brian M. D'Onofrio, Amber L. Singh, Anastasia Iliadou, Mats Lambe, Christina M. Hultman, Jenae M. Neiderhiser, Niklas Långström, and Paul Lichtenstein 80

Fetal Genotype for the Xenobiotic Metabolizing Enzyme *NQO1* Influences Intrauterine Growth Among Infants Whose Mothers Smoked During Pregnancy
Thomas S. Price, Tilo Grosser, Robert Plomin, and Sara R. Jaffee 101

Prenatal Antecedents of Newborn Neurological Maturation
Janet A. DiPietro, Katie T. Kivlighan, Kathleen A. Costigan, Suzanne E. Rubin, Dorothy E. Shiffler, Janice L. Henderson, and Joseph P. Pillion 115

Published by Wiley-Blackwell for the Society for Research in Child Development

Child Development

CHILD DEVELOPMENT (ISSN 0009-3920), one of three publications of the Society for Research in Child Development, Inc., is published bimonthly by Wiley Subscription Services, Inc., a Wiley Company, 111 River St., Hoboken, NJ 07030-5774. Periodical Postage Paid at Hoboken, NJ and additional offices.

Postmaster: Send all address changes to *CHILD DEVELOPMENT*, Journal Customer Services, John Wiley & Sons Inc., 350 Main Street, Malden, MA 02148-5020.

Information for Subscribers: *Child Development* is published in six issues per year. Institutional subscription prices for 2010 are: Print & Online: US$670 (US), US$938 (Rest of World), €608 (Europe), £479 (UK). Prices are exclusive of tax. Asia- Pacific GST, Canadian GST and European VAT will be applied at the appropriate rates. For more information on current tax rates, please go to www3.interscience.wiley.com/aboutus/journal_ordering_and_payment.html#Tax. The price includes online access to the current and all online back files to January 1st 1997, where available. For other pricing options, including access information and terms and conditions, please visit www.interscience.wiley.com/ journal-info.

Delivery Terms and Legal Title: Prices include delivery of print journals to the recipient's address. Delivery terms are Delivered Duty Unpaid (DDU); the recipient is responsible for paying any import duty or taxes. Legal title passes to the customer on despatch by our distributors.

For more information about Wiley-Blackwell journals, including online access information, terms and conditions, and other pricing options, please visit www.interscience.wiley.com or contact our customer service department, tel: (800) 835-6770 or +1 (781) 388-8598 (Americas); +44 (0) 1865 778315 (Europe, Middle East, and Africa); or +65 6511 8000 (Asia Pacific). Subscription in 2010 includes three issues of *Monographs of the Society for Research in Child Development* and three issues of *Child Development Perspectives.* Single issues from current and recent volumes are available at the current single issue price from csjournals@wiley.com. Earlier issues may be obtained from Periodicals Service Company, 11 Main Street, Germantown, NY 12526, USA. Tel: +1 518 537 4700, Fax: +1 518 537 5899, Email: psc@periodicals.com. Volumes 37 to date are available in microfilm from ProQuest Information and Learning, Serials Customer Service Department, 300 North Zeeb Road, Ann Arbor, MI 48106; telephone: (800) 521-0600 x2873.

Electronic Access: The *Child Development* website can be found at www.srcd.org. Abstract information for this journal is electronically available at www.blackwellpublishing.com/cdev. This journal is available online at Wiley InterScience. Visit www3.interscience.wiley.com to search the articles and register for table of contents e-mail alerts.

Access to this journal is available free online within institutions in the developing world through the HINARI initiative with the WHO. For information, visit www.healthinternetwork.org

Journal Customer Services: For ordering information, claims and any enquiry concerning your journal subscription, please go to interscience.wiley.com/support or contact your nearest office:

Americas: Email: cs-journals@wiley.com; Tel: +1 781 388 8598 or 1 800 835 6770 (Toll free in the USA & Canada).
Europe, Middle East, and Africa: Email: cs-journals@wiley.com; Tel: +44 (0) 1865 778315
Asia Pacific: Email: cs-journals@wiley.com; Tel: +65 6511 8000
Japan: For Japanese speaking support: Email: cs-japan@wiley.com; Tel: +65 6511 8010 or Tel (toll-free): 005 316 50 480.
Further Japanese customer support is also available at www.interscience.wiley.com/support
Visit our Online Customer Self-Help available in 6 languages at www.interscience.wiley.com/support

Membership communication and address changes should be addressed to the SRCD Executive Office, 2950 South State Street, Suite 401, Ann Arbor, MI 48104, USA. http://www.srcd.org/

Copyright and Photocopying: Journal Compilation © 2010 The Society for Research in Child Development, Inc. All rights reserved. No part of this publication (apart from articles marked "OnlineOpen") may be reproduced, stored or transmitted in any form or by any means without the prior permission in writing from the copyright holder. Authorization to photocopy items for internal and personal use is granted by the copyright holder for libraries and other users registered with their local Reproduction Rights Organization (RRO), e.g., Copyright Clearance Center (CCC), 222 Rosewood Drive, Danvers, MA 01923, USA (www.copyright.com), provided the appropriate fee is paid directly to the RRO. This consent does not extend to other kinds of copying such as copying for general distribution, for advertising and promotional purposes, for creating new collective works or for resale. OnlineOpen articles are made available in accordance with the terms of the Creative Commons Deed, Attribution 2.5 (further details available from www.creativecommons.org), which allows Open Access dissemination of the article, but does not permit commercial exploitation or the creation of derivative works. Special requests should be addressed to PermissionsUK@wiley.com.

Production Editor: Teresa Huang (email: CDEV@wiley.com).

Advertising: Kristin McCarthy (email: kmccarthy@wiley.com).

Indexing/Abstracting: The contents of this journal are indexed or abstracted in the following: *Academic ASAP; Expanded Academic ASAP; Academic Search Elite; Academic Search Premier; Applied Social Sciences Index and Abstracts; Biomedical Reference Collection: Comprehensive Edition; CAB Abstracts; CAB Direct; Contents Pages in Education; Current Contents/Social & Behavioral Science; Current Index to Journals in Education; Cambridge Scientific Abstracts Linguistics & Language Behavior Abstracts; CatchWord; Current Index to Journals in Education; Educational Research Abstracts; Environmental Sciences & Pollution Management; e-psyche; EBSCO Sociological Collection; EBSCO Online; EBSCO Professional Development Collection; Educational Research Abstracts; Educational Resources Information Center; Health and Safety Science Abstracts; ISI Basic Social Sciences Index; InfoTrac OneFile; Ingenta; JCR Social Sciences Edition; Leisure, Recreation, and Tourism Abstracts; Online Computer Library Center FirstSearch Electronic Collections Online; Periodical Abstracts Library; Periodical Abstracts Research II; ProQuest 5000; ProQuest Social Science Plus Text; PsycINFO; Psychological Abstracts; Psychology and Behavioral Sciences Collection; Pollution Abstracts; Safety Science & Risk Abstracts; Sage Family Studies Abstracts; Social Services Abstracts; Social Science Citation Index; Sociological Abstracts; Student Resource Center Bronze; Student Resource Center Gold; Student Resource Center Silver; Violence & Abuse Abstracts; Wilson OmniFile V; Wilson Social Sciences, Abstracts; and Wilson Social Sciences Index*

Published by Wiley-Blackwell for the Society for Research in Child Development

Child Development

Editor

Jeffrey J. Lockman
Tulane University

Associate Editors

Margaret Burchinal, *University of North Carolina at Chapel Hill*
John Colombo, *University of Kansas*
Michael Cunningham, *Tulane University*
Mary Dozier, *University of Delaware*
Andrew J. Fuligni, *University of California, Los Angeles*
Mary Gauvain, *University of California, Riverside*
Roberta M. Golinkoff, *University of Delaware*
Nancy G. Guerra, *University of California, Riverside*
D. Geoffrey Hall, *University of British Columbia*
Betsy Hoza, *University of Vermont*
Melanie Killen, *University of Maryland*
Louis J. Moses, *University of Oregon*
Cynthia Stifter, *The Pennsylvania State University*

Managing Editor

Detra N. Davis

Editorial Assistant

Antonella Caiazza

Consulting Editors

Emma Adam
Karen E. Adolph
Nazan Aksan
William F. Arsenio
Daniel H. Ashmead
Anna Axia
Karen Bartsch
Annie Bernier
Susan Birch
Amy Booth
Heather Bouchey
Paul Boxer
Gregory Stephen Braswell
Mara Brendgen
Gene H. Brody
Keith Burt
Emily W. Bushnell
Gustavo Carlo
Stephanie M. Carlson
Ruth K. Chao
Enrique Chaux
Pablo Chavajay
Tabbye Maria Chavous
Elysia Davis
David Dickinson
Gil Diesendruck
Ruth Feldman
Lisa Gatzke-Kopp
Tamsin C. German
Rick O. Gilmore
Artin Goncu
Susan A. Graham
Elena L. Grigorenko
Laura D. Hanish
Daniel Hart
David B. Henry
Claire H. Hughes
Mutsumi Imai
Deborah J. Johnson
Roger Kobak
Joachim Krueger
Richard Lee
Charlie Lewis
Jeffrey Lidz
Tina Malti
Jelani Mandara
Gary Marcus
Ann S. Masten
Bob McMurray
Julie Meeks-Gardner
Jayanthi Mistry
Ginger A. Moore
Dianna (Annie) Murray-Close
Velma McBride Murry
Letitia Naigles
Drew Nesdale
Nora S. Newcombe
David Oppenheim
Ana Papafragou
Susan M. Perez
Bradford Pillow
Eva Pomerantz
Douglas Powell
Kimberly K. Powlishta
Samuel Putnam
Suzanne M. Randolph
Robert Reeve
Kristen D. Ritchey
John J. Rieser
James L. Rodriguez
Glenn I. Roisman
Susan A. Rose
Martin D. Ruck
Jenny Saffran
Carlo Schuengel
David Schwartz
Judi G. Smetana
Alan L. Smith
Susan J. Spieker
Timothy Stickle
Dena Phillips Swanson
Adriana Umaña-Taylor
Christopher Trentacosta
Kimberly Updegraff
Marinus H. van IJzendoorn
Brian E. Vaughn
Elisa Velasquez
Lynne Vernon-Feagans
Peter Vishton
Elizabeth Votruba-Drzal
Cecilia Wainryb
Sarah Watamura
Henry M. Wellman
Mike Willoughby
Philip David Zelazo

Child Development, January/February 2010, Volume 81, Number 1, Pages 1–5

In This Issue

Crossing the Placenta

We used to think that fetuses were protected from external influences by the placenta, but we now know that characteristics of mothers' psychological state affect development of both the fetus and, after birth, the baby. Three studies in this issue address developments in this area.

Babies' development before they are born predicts their level of maturity in the first 2 weeks of life. In a longitudinal study, **DiPietro et al.** (p. 115) looked at about 100 healthy women with normal pregnancies when their fetuses were 24–38 weeks old. They measured fetal heart rate and how much the fetus moved to determine how this related to the maturity of the babies soon after birth. They found that more active fetuses had better motor development as newborns, and fetuses with more variation in their heart rate had more mature nervous systems. Additionally, they found that pregnant women who felt stressed by their pregnancies also had infants with more mature nervous systems than those who felt less stressed. The study broadens our understanding of how early development carries over to later development.

Davis and Sandman (p. 131), in a longitudinal study looking at 125 women and their babies, considered how prenatal exposure to maternal cortisol (a hormone produced in response to stress) during pregnancy affects infants' mental development. Infants who were exposed to high levels of cortisol early in pregnancy and whose moms were anxious about pregnancy performed poorly on measures of mental development at age 1. But infants who were exposed to high levels of cortisol late in pregnancy had enhanced mental abilities. In terms of the influence of mothers' cortisol on developing nervous systems, the study suggests that the timing of exposure matters and that fetuses are most vulnerable to the negative effects of stress in early pregnancy. The study can inform efforts to promote infant and child health.

Children whose mothers suffer from depression are more likely than others to develop behavioral and emotional problems. In a longitudinal study, **Hay, Pawlby, Waters, Perra, and Sharp** (p. 149) considered the role of mothers' depression during pregnancy by looking at more than 100 British youth from disadvantaged areas. Mothers who became depressed when pregnant were four times as likely to have children who were violent at 16—this was true for both boys and girls. Furthermore, women who were aggressive and disruptive in their own teen years were more likely to become depressed in pregnancy, so that the moms' history predicted their own children's violence. The findings encourage providing support for moms with a history of conduct problems who become depressed during pregnancy.

Autism: Exploring Cause and Effect

Research has shown that moms of children who develop autism have more problems during pregnancy and that their children have more problems during and immediately after birth. Do these early problems increase a child's risk of developing autism through an environmental mechanism, or are the pregnancy and birth problems themselves a consequence of the child's preexisting genetic risk for autism? In a longitudinal study, **Ronald, Happé, Dworzynski, Bolton, and Plomin** (p. 166) looked at more than 13,000 18-month-old identical and nonidentical twins born in England and Wales to see if the twin with more postbirth problems developed more autistic features by age 8 than the cotwin with fewer problems. The conclusion: a weak trend toward twins with more postbirth difficulties having more autistic features in middle childhood. Further research is needed to confirm and expand upon these findings, which have implications for treatment during pregnancy.

Early Risks and Later Outcomes

Some 4.2 million infants, toddlers, and preschoolers lived in poverty in the United States in 2007. We know that their childhoods are more difficult than those of affluent children, but what does low income in childhood mean for successes and failures later in life? **Duncan, Ziol-Guest, and Kalil**

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8101-0001

(p. 306) used nationally representative data of more than 1,000 people born between 1968 and 1975 (the Panel Study of Income Dynamics, the longest running longitudinal study of household income in the United States) to find strong links between poverty before age 5 and lower earnings and fewer work hours 30 years later. In concluding that the age at which children live in poverty matters, the study has implications for policy.

Children who experience many early risk factors—living in poverty, being born to a teen mom, moving frequently—are more likely to have problems doing well in school and adjusting socially. In a longitudinal study of more than 100 low-income mostly African American children from infancy to age 21, **Pungello et al.** (p. 410) looked at early risk and early intervention together. They found that children who experienced more early risk were less likely to graduate from high school and get a job, and more likely to become parents as teens, whereas children who took part in an early intervention preschool program from infancy to kindergarten were more likely to attend college or get a good job. In suggesting that the intervention buffered against the harmful effects on education of living in a less than optimal home environment, the findings can help policy makers decide how to allocate limited resources.

Watching TV

American children watch a lot of TV—3.6 hours a day by the time they are 3—and there has long been concern about how TV affects children's development. In a reanalysis using data from the National Longitudinal Survey of Youth, **Foster and Watkins** (p. 368) reexamined the claim made by prior studies that TV watching in the toddler years causes attention deficit disorder (ADD) at age 7. They find no meaningful tie between TV watching and attention problems. Prior research claiming the existence of a TV–ADD link, the authors suggest, omitted key confounding variables such as poverty. This study broadens our understanding of the effects of TV on children, and also highlights the value of reanalyzing past research.

Maternal Smoking During Pregnancy—and Its Effect on Children

Mothers who smoke while pregnant tend to give birth to smaller, less healthy babies, while mothers who quit smoking are more likely to have babies with normal birth weight. This tells us that exposure to tobacco smoke harms fetal development, but less is known about the physiological processes that cause this to happen. **Price, Grosser, Plomin, and Jaffee** (p. 101) looked at the DNA of almost 500 nonidentical twins born to moms in England and Wales who smoked during pregnancy—specifically, the gene that codes for a protein that forms part of the body's defenses against the toxic chemicals found in tobacco smoke. Differences in this gene were associated with differences in birth weight. In twin pairs where one child had an inactive form of the gene and the other did not, the twin with the inactive gene tended to be born several ounces heavier. Identifying this gene provides a keyhole for future experiments on the biology of fetal development when there is exposure to smoke.

Children of mothers who smoke during pregnancy are also at risk of doing poorly academically. In a study that is based on information from more than 650,000 Swedish children and that used a quasi-experimental approach, **D'Onofrio et al.** (p. 80) suggest that it's the risk factors that accompany mothers' smoking, including shared genetic factors, that lead to poor academic achievement (measured by school grades and math proficiency at age 15), not the smoking itself. In challenging prior findings, the study has implications for public health interventions.

Caring for Children in Institutions

Research has shown that institutional care may hinder both emotional and intellectual development in young children. Based on data from the Bucharest Early Intervention Project, the first randomized controlled trial of foster care as an alternative to institutional care, **Smyke, Zeanah, Fox, Nelson, and Guthrie** (p. 212) examined whether a family-based intervention could improve the quality of attachments between Romanian 3-year-olds and their caregivers. They studied children who had been in institutions, usually from birth; children institutionalized at birth and later placed in foster care; and children raised by their families at home. Young institutionalized children who were placed in foster care before 24 months were much more likely to develop secure attachments to their caregivers than those placed after 24 months. Children cared for in institutions were more likely to have unusual and abnormal attachments to their caregivers. The study has implications for abandoned and abused or neglected children around the world who receive nonparental care, as well as for their ability to recover from their early experiences.

How does the deprivation and neglect that institutionalized children often experience affect their brains' development? **Pollak et al.** (p. 224) looked at groups of 8- and 9-year-olds, some of whom were adopted into U.S. homes from institutions in Asia, Latin America, Russia, Eastern Europe, and Africa. These children were compared to a group of American children raised in their birth families. Children who lived much of their preadoptive lives in institutions performed worse than those raised in families on tests measuring visual memory and attention, learning visual information, and impulse control. Yet these children performed at developmentally appropriate levels on similar tests involving attention, memory, and learning sounds, such as sentences, words, and phonemes. Both groups of children also performed similarly on tests of planning abilities. The take-home message: The amount of time young people spend in deprived settings may have different effects on the development of brain systems that underlie how they learn different types of information, such as sights versus sounds.

Studies have suggested that institutional care jeopardizes children's optimal development and that well-functioning child-care institutions may offer children a better environment than their own dysfunctional families. For the growing number of children who are becoming infected with HIV, questions about where they live and who cares for them are crucial. **Dobrova-Krol, van IJzendoorn, Bakermans-Kranenburg, and Juffer** (p. 237) examined the effects of perinatal HIV infection on the development of more than 60 Ukrainian 4-year-olds; some lived with their biological families and some lived in child-care institutions from shortly after birth. The quality of the child–caregiver relationship had a bigger impact on children's physical growth and cognitive performance than that of the HIV infection or the quality of the physical environment, they find. In addition, for both children with and without HIV who lived in institutions, family care, even when it was compromised, was better for children than institutional care. The study has implications for intervention efforts among children with HIV.

Effects of Early Events on Vulnerable Children

Children who are abused and neglected grow up with a lot of stress. Cortisol, termed the "stress hormone," helps the body regulate stress. But when stress is chronic and overloads the system, cortisol can soar to very high levels or plummet to lows, which in turn can harm development and health. **Cicchetti, Rogosch, Howe, and Toth** (p. 252) studied more than 500 low-income children ages 7–13, about half of whom had been neglected and/or abused, to find out whether abuse early in life and feelings of depression affected their levels of cortisol. The study's findings: Children abused in the first 5 years had more depression than maltreated children without early abuse and children who were not maltreated. In addition, only children who were abused in the first 5 years and were depressed had an atypical flattening of cortisol production during the day. The authors suggest that early abuse may be more damaging to developing emotion and stress symptoms because it happens as the brain is rapidly developing. Moreover, because it's harder for very young children to discern the clues predicting an abusive attack, they may be chronically stressed and overly vigilant, even when they are not being abused.

Early events in our lives influence how both our brains and our behavior develop. **Fox, Levitt, and Nelson** (p. 28), in a review of past research, explore how the structure of early experience weaves its way into the structure of the developing brain. First, they describe the genetic framework that lays the foundation for brain development; then, they turn to the many ways experience interacts with and modifies the structures and functions of the developing brain. Their conclusions: Although early experience is key to maintaining and elaborating our early wiring diagram, later experience also plays an important role; this later experience is critical to establishing a solid footing for development beyond the early years. The analysis has implications for policy and practice for those who work with vulnerable children.

Thought Processes and Processing Thoughts

Executive functioning is a set of advanced cognitive functions—including impulse control, working memory, and mental flexibility—that help us plan and monitor what we do to reach goals. Although executive functioning develops speedily between ages 1 and 6, children vary widely in their skills in this area. In a longitudinal look at 80 pairs of middle-income Canadian moms and their year-old babies, **Bernier, Carlson, and Whipple** (p. 326) asked why. It turns out that the ways moms act when they are playing and solving puzzles with their babies can explain some of the differences. Children of moms who answered their children's requests for help quickly and accurately; talked about their children's preferences, thoughts, and memories during play; and used successful strategies to help solve difficult problems performed better at a year and a half and 2 years on tasks that call for executive skills. The study sheds light on the role parents play in helping children develop skills deemed important for later school success and social competence.

In everyday life, people often reflect on how a past event might have turned out differently—a process called counterfactual reasoning. Some studies suggest that even 3-year-olds can reason counterfactually. However, in studies assessing feelings of regret, children show no sign of counterfactual reasoning until age 6 or older. To explore why this is so, **Rafetseder, Cristi-Vargas, and Perner** (p. 376), in a series of experiments, studied 3- to 6-year-old Austrian children as well as adults, asking them to answer counterfactual questions. Some of the questions could not be answered correctly by basic conditional reasoning ("if . . . then"), but required the respondents to think of what had actually happened. The study finds that counterfactual problems become quite difficult when basic conditional reasoning leads to a different answer than counterfactual reasoning does. So the early use of basic conditional reasoning explains why, in some studies, little sign of counterfactual reasoning emerges before age 6.

Biology Versus Psychology

Some research suggests that young children don't know a lot about biology and so they say that biological occurrences are caused by psychological factors—for example, a sibling is shaking because of fear, not a fever. **Erickson, Keil, and Lockhart** (p. 390) used implicit measures to examine whether children can distinguish biological from psychological phenomena, and how they reason differently about the processes in these two distinct categories. They conclude: Five-year-olds correctly cluster biological processes and separate them from psychological ones. Also, 5-year-olds understand that biological and psychological processes operate according to fundamentally different causal mechanisms. And they have intuitions about the broad nature of these different mechanisms—even though they may not understand the specifics behind them. This study expands our understanding of how and when children reason about biological happenings as separate from psychological ones.

Studies of Stress

Children who react highly to stress are more vulnerable to adversity and have more behavior and health problems than their peers. But a growing body of research tells us that this picture is incomplete. **Obradović, Bush, Stamperdahl, Adler, and Boyce** (p. 270) took a longitudinal look at more than 300 kindergarteners, as well as their teachers and families. Children who had significant biological reactions to laboratory challenges that mimicked daily stressors were more susceptible not only to negative environments but also to positive environments. Children who were highly reactive were more likely to have developmental problems when growing up in adverse, stressful family settings. But contrary to expectation, such children were also more likely than their less reactive peers to thrive in caring, low-stress families because of their sensitivities to the supportive and nurturing qualities of such environments. The study tells us that when children are highly susceptible to stress, it's not always bad news, but rather should be considered in terms of the type of environment they live in.

In another study on the effects of early experience on children's reactions to stress, **O'Neal et al.** (p. 290) sought to understand the relation between children's response to stress (as measured by cortisol) and aggressive behavior in almost 100 preschoolers who were at risk of antisocial behavior (their older siblings had been in trouble with the law). The study finds that in families with lower levels of parental warmth, changes in the children's response to stress that came about as a result of a preventive intervention program were associated with lower rates of aggressive behavior 16 months later. This is one of the first experimental studies with young children to establish a tie between the neurobiological stress regulation system and later aggression. By underscoring the malleability of the neurobiological system and links to behavior, the study has implications for those who work with children living in stressful circumstances.

Because the survival rate of babies born at extremely low birth weight (less than 1,000 g or 2.2 pounds) has improved markedly over the last few decades due to advances in neonatal care, it's important to track how these babies fare. In a longitudinal study, **Schmidt, Miskovic, Boyle, and Saigal** (p. 183) looked at the differences between

Canadian children born at extremely low birth weight and those born at normal birth weight in terms of anxiety, depression, and social withdrawal in middle childhood, adolescence, and young adulthood; they also considered brain-based measures of vulnerability to stress. Adults who were born at extremely low birth weight had more behavior problems from adolescence to young adulthood and their brains were more vulnerable to stress. The findings shed new light on the development of those born at extremely low birth weight, especially in terms of the ties between brain and behavior.

Refining the Nature–Nurture Debate

Children's behavior is shaped by genes and environment. Studies suggest that aggressive and antisocial behavior is influenced by moderate genetic effects, and environmental factors moderate or mediate the genetic influences on problem behaviors. Expanding work on how inherited characteristics make some more susceptible to the effects of certain environments, **Leve et al.** (p. 340) looked at infants' reactions to frustration. In their study, part of ongoing longitudinal research, the scientists observed almost 350 nine-month-olds and their adoptive parents during a frustrating task; they also interviewed birth parents. Perhaps not surprisingly, children were more likely to engage in problem behaviors on the task when they had adoptive mothers who were anxious or depressed *and* biological moms with a history of problems with substance abuse, delinquency, or other issues. But encouragingly, the adoptive children whose biological moms had problems and whose adoptive moms were *not* anxious or depressed did not act out on the frustrating task. The findings support the idea that healthy family environments can offset a child's genetic risk for difficulties, even during the 1st year of life.

Individual differences in human traits commonly flow from parent to child, prompting the question: Do children share the traits of their parents because they are raised by their parents or because they share common genes? **Meaney** (p. 41) reviewed research to examine the biological context in which genetic and environmental signals operate. He suggests that the nature–nurture debate may appear sensible, but it's really a false debate because when we understand how genes actually work within the cells of our body, it becomes clear that genetic forces simply cannot operate independently of environmental forces. The bottom line: Genetic and environmental influences—nature and nurture—work collaboratively, not as separate and opposing forces, to shape children's development. The analysis has implications for prevention and intervention programs for children and families.

Reversing Intellectual Disability in PKU Children

Phenylketonuria (PKU) is an inborn metabolic error caused by a genetic mutation that results in higher levels of phenylalanine, which can cause serious brain damage. Newborn screening for PKU, begun in the 1960s, combined with a special diet has been found to prevent intellectual disability in children with PKU. It has been thought that when children with PKU are not treated or are treated late, they suffer severe, irreversible mental retardation that cannot be reversed through diet. **Grosse** (p. 200) reviewed studies from Europe and North America and finds that following a restricted diet can boost cognitive ability in those with late-diagnosed PKU. This conclusion should encourage parents of children who missed being detected through newborn screening. That said, the study also reaffirms that early diagnosis and intensive treatment are the only ways to minimize neurological damage caused by PKU, so universal screening and prompt treatment remain urgent public health issues.

Anne Bridgman

Child Development, January/February 2010, Volume 81, Number 1, Pages 6–22

A Unified Theory of Development: A Dialectic Integration of Nature and Nurture

Arnold Sameroff
University of Michigan

The understanding of nature and nurture within developmental science has evolved with alternating ascendance of one or the other as primary explanations for individual differences in life course trajectories of success or failure. A dialectical perspective emphasizing the interconnectedness of individual and context is suggested to interpret the evolution of developmental science in similar terms to those necessary to explain the development of individual children. A unified theory of development is proposed to integrate personal change, context, regulation, and representational models of development.

The attention of philosophers and then scientists to human development has always begun with a concern that children should grow up to be good citizens who would contribute to society through diligent labor, moral family life, civil obedience, and, more recently, to be happy while making these contributions. The motivation for these concerns was that there were many adults who were not. Although attention was paid to the socialization and education of children, it was ultimately in the service of improving adult performance. The societal concern has always had a life-span perspective. Without healthy, productive adults no culture could continue to be successful. This concern continues to be a major motivator for society to support child development research. Although the intellectual interests of contemporary developmental researchers range widely in cognitive and social–emotional domains, the political justification for supporting such studies is that they will lead to the understanding and ultimate prevention of behavioral problems that are costly to society.

With these motivations and supports there have been major advances in our understanding of the intellectual, emotional, and social behavior of children, adolescents, and adults. Moreover these understandings have increasingly involved multilevel processes cutting across disciplinary boundaries in the social and natural sciences. This progress has forced conceptual reorientations as earlier unidirectional views that biological or social circumstance controlled individual behavior are becoming multidirectional perspectives where individual behavior reciprocally changes both biological and social circumstance.

The models we use to understand how individuals change over time have increased in complexity from linear to interactive to transactive to multilevel dynamic systems. Was this progression in complexity an expression of empirical advances in our developmental research or is it related to more general progressions in the history of science as a whole? Several years ago during a discussion of a need for a critical social history of developmental psychology by a number of distinguished scientists (Bronfenbrenner, Kessel, Kessen, & White, 1986), Sheldon White argued that it is necessary to engage and deconstruct the history of the field in parallel with efforts to understand the child. He continued by pointing out that the study of development needs a self-concept, just as each child requires ''the building of some kind of self-referential, self-regulating, self-knowing set of structures.''

If there is a more sophisticated understanding of the development of humans, is there a more sophisticated understanding of the development of our science? The models we use to understand the history of our field from child psychology to developmental science should increase in complexity. Understanding developmental science requires developmental science. And as in the study of any

---

Correspondence concerning this article should be addressed to Arnold Sameroff, Center for Human Growth and Development, University of Michigan, 300 N. Ingalls Building, Ann Arbor, MI 48109-0406. Electronic mail may be sent to sameroff@umich.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8101-0002

historical process there should be hope that understanding the past will help us predict the future.

The premise of the general systems theories that arose in the 1930s was that there were general principles of organization in every scientific domain that were at a level of abstraction somewhere between mathematical formulations and the specific processes being studied (Boulding, 1956). This has become apparent in every discipline from physics to political science, as each has moved to models of dynamic regulation, where parts cannot be separated from wholes and useful predictions can only be made based on local interactions of multiple systems. The hope of the founders of general systems theory (cf. von Bertalanffy, 1968) was that scientists would use a top-down strategy to interpret empirical data from a complexity perspective (Sameroff, 1983). This aspiration was not realized because each science has tried to be as theoretically simplistic as possible, resisting the demise of deterministic models until overwhelmed by the complexity of empirical data. The science of psychology has been no exception.

Developmental research aspired to the dicta of Ockham's razor in the hope of finding simple basic elements and processes that would explain the emergence of life's complexity. Up through the 1960s and into the 1970s statistically significant *t* tests and analyses of variance gave an illusion that science was advancing, but when regression models became dominant and the metric changed to size of effects (Cohen, 1988), it became clear that the field was not doing well at explaining how children were growing up. Contemporary developmentalists are quite competent at short-term predictions of similar cognitive or emotional constructs but much worse at the prediction of long-term successful life adaptations starting from initial conditions. Increasingly, sophisticated statistical models have been sought to separate the behavioral signal of interest from the noise of real life. This effort has led to some frustration in the decreasing amounts of variance that can be attributed to any single factor when everything imaginable is controlled and obscured the possibility that the unexplained variance, the noise, might contain the signals of many other dimensions of the individual or context that are necessary for meaningful long-term predictive models.

Applicability may not be the most salient criteria for getting research accepted for publication, but it is highly salient for suggesting ways to change developmental outcomes. The science paid for by the public is increasingly being asked to meet a translational rather than a statistical criterion with the application of research to policy an important consideration (Huston, 2008). The primary question remains as to how we can improve the fate of individuals growing up in our society. To answer that question requires a continuing examination of the models we need both to study and to understand development. In what follows I will present a contemporary summary of what such models should contain and offer a suggestion for an integrated view of development that captures much of the variance that needs explaining. No part of what I propose has not been previously suggested by creative others. Combining these elements into a unified developmental theory acknowledges the contemporary zeitgeist moving toward more dynamic conceptualizations at every level of analysis that is taking place in every other scientific discipline.

A Rough History of the Nature Versus Nurture Question

Before complexity was simplicity. For developmental explanations, simplicity was expressed in appeals to aspects of an individual's nature or nurture. The history of developmental psychology has been characterized by swings between opinions that determinants of an individual's behavior could be found either in their irreducible fundamental units or in their irreducible fundamental experiences. The growth process between babyhood and adulthood could be explained either by appeals to *intrinsic* properties of the child or to *extrinsic* properties of experience. The nature–nurture question has been a central *content* of developmental research, but it can also be considered to be a major *context* for developmental research in its appeal to deterministic thinking. As a consequence the history of the nature–nurture question can be used as an organizing construct to understand the history of our field.

Practically, the nature–nurture question comes into play when a child has a problem and the question arises, "Who is responsible?" Most parents' first response is to blame the child and most professionals' first response is to blame the parents. However, most scientists know that it is both. It is both child and parent, but it is also neurons and neighborhoods, synapses and schools, proteins and peers, and genes and governments. But that conclusion does not explain how it is both. Do nature and nurture interact deterministically so that the proportions attributable to each can be decomposed or do they transact probabilistically so that the contribution of each can only be an abstraction from the activity of dynamic systems? How this question has been answered in the course of recent history offers a window into how developmental science has evolved and a perspective on how the question will be answered in the future.

Since ancient times philosophers have weighed in with their perspectives on the relative influences of constitution and experience in determining the life course, but it is in the last few hundred years that these positions have been well articulated, most notably John Locke in the 17th century and Rousseau in the 18th. I will begin my rough historical account in the late 19th century with the beginnings of empirical psychological research in the work of Francis Galton (see Table 1). Francis Galton coined the "nature versus nurture" phrase and in his view inherited characteristics were the origins of human nature. The nurture counterpoint was most strongly stated in the work of John Watson in the 1920s who propounded a new approach he labeled *behaviorism*, extending Pavlov's conditioning processes to explain human individual differences. Learning theory came to dominate human developmental research for almost 50 years strengthened by the operant paradigms promoted in the work of the Skinnerians.

This tilt toward nurture began to shift in the 1960s under assault from three directions—ethology, behavioral genetics, and the cognitive revolution. Where S-R theorists had argued that the laws of learning were primary in explaining developmental change, ethologists were demonstrating that many complex behaviors did not seem to need any reinforcement (Lorenz, 1950) and that S-R contingencies that worked in one species did not work in another (Breland & Breland, 1961). For example, rats could learn to push a lever to avoid a shock but pigeons could not. Ethologists argued that the nature of the species put large restrictions on the effects of nurture such that certain prepared responses were impervious to experience (Seligman, 1970). Statistical advances and data from large samples of twins permitted behavioral geneticists to argue that the effects of genes and environments could be separated, and that very large proportions of behavioral differences could be explained by genetic differences (Defries & McLearn, 1973). The cognitive revolution characterized in the work of Jean Piaget placed the source of development in the mind of the child. Experience was necessary for the child to construct the world but it did not play a role in individual differences.

Table 1
Rough History of Nature–Nurture

Historical era	Empirical advance
1880–1940s—Nature	Inherited differences Instincts
1920–1950s—Nurture	Reinforcement theory Psychoanalytic theory
1960–1970s—Nature	Ethology—species differences Behavioral genetics Cognitive revolution
1980–1990s—Nurture	Poverty Social ecology Cultural deconstruction
2000–2010s—Nature	Molecular biology Neuroscience

Where the nativist shift in the 1960s was driven by advances in biological science, the nurturist shift in the 1980s was driven by three advances in the social science—the war on poverty, the concept of a social ecology, and cultural deconstruction. Where behaviorist research focused on proximal connections between reinforcements and performance, scientists in other social disciplines were arguing that economic circumstance was a major constraint on the availability of reinforcements, such that the developmental environments of the poor were deprived in contrast with those of the affluent. Similar individuals in different social classes would have quite different developmental outcomes. Bronfenbrenner (1977) in his vision of the social ecology offered a more differentiated model than provided by economics alone. He identified the distal influences of family, school, work, and culture on the availability of reinforcements to the child, providing a more comprehensive empirical model for predicting individual differences in development. The influence of postmodernist deconstruction was manifest in the emergence of a cultural psychology that went beyond cross-cultural descriptive studies. Meaning rather than behavior became dominant through demonstrations that the same child behaviors could be given different meanings in different societies leading to different developmental consequences, and conversely, different behaviors could be given the same meaning leading to the same consequences.

The new millennium coincided with another swing of the pendulum in the nativist direction, again tied to major advances in biological science. Neuroscience and molecular biology have been making major contributions to our understanding of development with new technologies for imaging the brain and manipulating the genome. But, as will be discussed below, the more recent swings between nature and nurture have been getting shorter and their intermingling has been increasing.

An examination of Table 1 emphasizes the swings between the popularity of nature and nurture as developmental explanations. At each point in time there are strong adherents of both positions waiting for some new technological advance to reinforce their point of view. Although this polarity provides motivation for empirical innovation, it has the unfortunate side effect of inhibiting theoretical innovation. Despite the alternating claims that the argument is now closed by those on the frontier of new explorations of nature or nurture, the fact remains that after each advance most of the variance in long-term developmental outcomes is still unexplained. It is the pressure of unexplained variance that continually negates claims of ascendancy and dialectically motivates continuing exploration.

I have presented a descriptive case for the cycling of explanations between nature and nurture to raise the question if there is an explanation of the repetitive pattern. It could be interpreted as simply the result of technological or theoretical advances, but it also could be a phenomenon in itself. The development of the nature–nurture debate might follow developmental principles similar to those that regulate human development and the examination of the two in parallel might illuminate both.

Nonlinear Models of Development

An appreciation of cycling requires an appreciation of a number of nonlinear processes that I will discuss under the general rubric of dialectical theory with specific attention to a developmental helix and processes of differentiation and integration. Dialectics have been directly or indirectly emphasized for studying development and especially relationships (Hinde, 1997; Riegel, 1976). An initial approach to dialectics is best captured by consideration of the Taoist diagram of the dark yin and the light yang (see Figure 1) that emphasizes that opposites are in a mutually constituting relationship. They were created together and remain bound to each other. This philosophical statement is empirically validated at the most fundamental level of physics where quarks, the current basic entities, are always in a relationship with each other. At the most fundamental level of the universe there are no ultimate units, only ultimate relationships. In the dialectical yin–yang there is a *unity of opposites* and an *interpenetration of opposites*. The unity is indicated by the mutual embrace of the yin and the yang, as seen in the figure, but yin and yang also interpenetrate each other as depicted by the small black spot of yin within the yang and small white spot of yang within the yin.

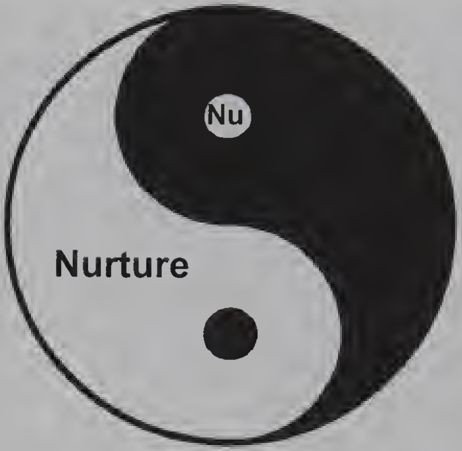


Figure 1. Unity of opposites and interpenetration of opposites in yin and yang diagram. Nu = nurture; Na = nature.

In the psychological realm these ideas have been applied frequently, beginning with the philosophical writings of Hegel and most manifest in Piaget's theory of cognitive development. There is a unity of opposites between one's cognitions and the world that is being cognized. Without the world there would be nothing to cognize, and without the cognizer there would be no cognitions. But there is also an interpenetration of opposites. One's cognition leads to one's action which becomes part of the world (the small black dot in the white area), and then the changed world becomes a part of one's cognition (the small white dot in the black area) in a continuing dialectical progression.

The dialectical perspective on nature and nurture is that they mutually constitute each other. There is a *unity of opposites* in that development will not occur without both, and there is an *interpenetration of opposites* in that one's nature changes one's nurture and conversely one's nurture changes one's nature, as captured in current transactional models. Moreover, and most salient, without the one, the other would not exist. Species and their environments evolved together in a coactive and transactional relationship. Gottlieb's (1992) construct of probablistic epigenesis centered on the joint regulation by organismic and experiential factors that produced development with neither having priority over the other. The reciprocal bootstrapping between cultural change in groups and cognitive change in individuals is well articulated by Cole (2006) in his description of human phylogeny.

Although Galton and Watson are the straw men that nurturists and nativists, respectively, rail against, both appreciated the unity of constitutions and environments. Galton (1876) recognized the influence of social class and wrote, "Nature prevails enormously over nurture when the differences of nurture do not exceed what is commonly to be found among persons of the same rank in society and in the same country." Watson (1914), in turn, recognized that individual and species differences were important, "effectiveness of habit training would be facilitated by knowledge of an animal's individual instinctive responses." The unity and interpenetration of nature and nurture will be more fully explored in the unified model of development to follow.

The Developmental Double Helix

The dynamic dialectical interplay between opposites can best be captured as an image of a helix that depicts the developmental aspects of changes over time as can be seen in Figure 2a. A simple example of a developmental progression is the daily cycle where spiraling to the right would be the movement toward day and spiraling left would be the movement toward night. Although this is a repetitive cycle, it becomes helical in that each day is different because of the experience of the previous night and each night is different because of the experience of the preceding day. A more complex example would be the development of representation in children (Werner, 1948). Initially, infants represent the world as images of here and now experiences. Preschoolers cycle over the same material but now have the capacity to depict images in drawings that may have a one-to-one correspondence to the images but are not the same as the images. In a few years they will recycle over the same contents but now with the ability to do abstract representations such as maps where the pictorial aspects may be completely eliminated in favor of words and symbols. Such developmental recycling also occurs in the social-emotional domain where relationship experiences and representations derived from early parent–child relationships are reworked as children enter into peer relationships and reworked again in the romantic relationships beginning in adolescence. Erikson (1959), although not known for his empiricism, was very articulate in describing the recycling of identity issues that are never resolved but through a balancing of opposites provide the impetus for each succeeding stage. The figure of the helix emphasizes that the same issues in a variety of domains are revisited again and again during development. The ubiquity of this helical concept is even found in Graduate Record Examination practice questions

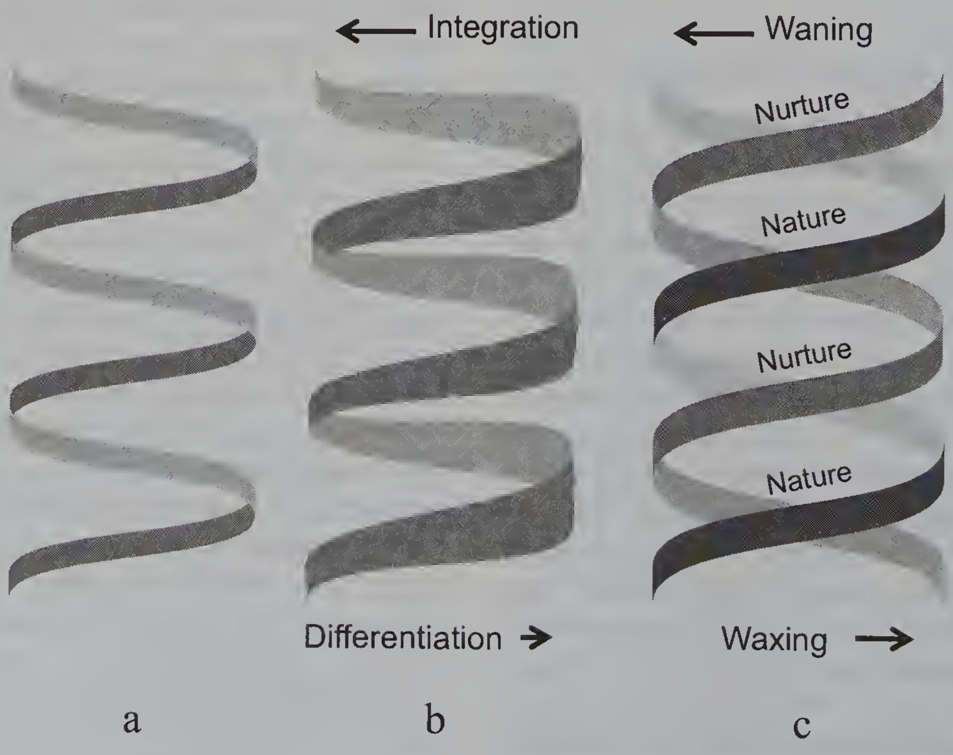


Figure 2. (a) Developmental helix. (b) Differentiation and integration of helix. (c) Developmental double helix of nature and nurture.

(Princeton Review, 2009) where a correct answer is "Science advances in a *widening* spiral in that each new conceptual scheme *embraces* the phenomena explained by its predecessors and adds to those explanations."

Differentiation and Hierarchic Integration

The developmental helix pushes us toward a more elaborate nonlinear process expressed as differentiation and hierarchic integration. As formulated in Werner's (1957) orthogenetic principle, "Wherever development occurs it proceeds from a state of relative *globality and lack of differentiation* to a state of *increasing differentiation, articulation, and hierarchic integration.*" If viewed within the helical metaphor in Figure 2b, we could consider the movement toward differentiation as going in one direction with a widening of the coil, as for example, the number of words in a child's vocabulary or the number of color concepts increases, and then the movement toward integration going in the other direction with a narrowing of the coil, as for example, the chunking of metacognition occurs, only once again to begin differentiating again as the number of metaconcepts increases.

If we consider the historical differentiation of nature, what began in Galton's laboratory as a catalogue of measurable differences in behavior was reconceptualized as really being differences in neurological electrical activity, and then as really being differences in neurotransmitter activity, and then as really being differences in genomic activity, and most recently as really being differences in epigenomic activity.

Analogously, there was also a historical differentiation of nurture where an early romantic conception of the power of mother love was reconceptualized as differences in the pattern of reinforcements provided by the parent, and then reconceptualized when it was discovered that differences in social circumstance constrained the patterns of reinforcement available to the child, and then reconceptualized when social circumstance was differentiated into the subsystems of the child's social ecology, and then reconceptualized when it was realized through social deconstruction that the effects of social ecology were constrained by the meanings that families and cultures imposed on behavior.

The progression of nature and nurture conceptions can be summarized by a double helix that captures their alternating differentiation and integration waxing and waning through time (see Figure 2c). Each new breakthrough initially goes through a stage of differentiation as a new methodology comes into play and then integration as it becomes connected to developmental phenomena. The developments in molecular biology would be a recent example on the nature side where the genome project produced the differentiated genes that now can be integrated into endophenotypes that have more proximal connections to behavior. On the nurture side, the differentiation of the social ecology into a set of subsystems of family, school, peer group, and neighborhood influences, for example, led to efforts at integrating its effect on development within comprehensive statistical models. Whether one gains ascendance over the other is a complex result of psychology (e.g., it is easier to conceptualize the parts we are made of than the wholes of which we are parts), anthropology (e.g., the preference in Western culture for individual-based rather than relationship-based explanations of behavior), sociology (e.g., whether there is a greater societal demand to mitigate the effects of biological disease or social disorder), and economics (e.g., whether investments in nature or nurture research offer the best opportunity to reduce the costs of developmental problems).

What is important in this discussion is to appreciate that there is a cycling between nature and nurture explanations of development that have a developmental course. The development of our science may be very similar to, and thus very useful for, understanding the development of human beings. The dialectics of differentiation and hierarchic integration may characterize all developmental processes.

We can come away from this discussion with one of two propositions. The first is that the cycling between nature and nurture will continue until either one or the other gets it right effectively ending the argument. Unfortunately, the problem of multifinality and equifinality undercuts this possibility (Cicchetti & Rogosch, 1996). On the nature side, whatever measure of individual differences has been discovered, two children with the same characteristics can have quite different outcomes and two children with different characteristics can have the same outcome. On the nurture side, whatever measure of the social environmental has been discovered, two children with the same experiences can have different outcomes and two children with quite different experiences can have the same outcome.

The second proposition is that nature and nurture represent a unity of opposites such that neither can ever get it right on its own. Because of their interpenetration advances in our understanding of nature illuminate nurture and changes in our understanding of nature illuminate nature. Although the literature contains many references to the fact that one cannot separate nature from nurture, there are fewer references to how their unity operates. The rest of this article will be devoted to an effort to integrate contemporary advances in our understandings of both nature and nurture into a unified theory of development.

A Unified Theory of Development

Contemporary developmental science requires at least four models for understanding human growth: a *personal change* one, a *contextual* one, a *regulation* one, and a *representational* one. The *personal change model* is necessary for understanding the progression of competencies from infancy on. It requires unpacking the changing complexity of the individual as he or she moves from the sensorimotor functioning of infancy to increasingly intricate levels of cognition; from early attachments with a few caregivers to relationships with many peers, teachers, and others in the world beyond home and school; and from the early differentiation of self and other to the multifaceted personal and cultural identities of adolescence and adulthood. The *contextual model* is necessary to delineate the multiple sources of experience that augment or constrain individual development. The growing child is increasingly involved with a variety of social settings and institutions that have direct or indirect impact as exemplified in Bronfenbrenner's (1977) view of the social ecology. The *regulation model* adds a dynamic systems perspective to the relation between person and context. During early development, human regulation moves from the primarily biological to the psychological and social. What begins as the regulation of temperature, hunger, and arousal soon turns to regulation of attention, behavior, and social interactions. The last is the *representational model* where an individual's here and now experiences in the world is given a timeless existence in thought. These representations are the cognitive structures where experience is encoded at abstracted levels that provide an interpretive structure for new experiences, as well as a sense of self and other. Combining these four models offers a comprehensive view of the multiple parts, wholes, and their connecting processes that comprise human development.

Personal Change Model

Because psychology's central focus is on individuals, developmental psychology's main concerns have been on how children change over time. How one thinks about developmental change will have a clear influence on research objectives. Three ways of conceptualizing change can be seen in Figure 3—trait, growth, and developmental. If one believes that an individual consists of a set of unchanging traits then there is no need for developmental research. If development is considered a growth process then it can have classic epigenetic explanations in that all the parts are there to start with and it is their interactions that produce the changes in the phenotype, or it can be considered experience dependent but only as nutrition for the unfolding maturation process. Viewing personal change as a stage process can have a descriptive or theoretical meaning (Kessen, 1962). Descriptive stages are paraphrases for age and consist of lists of average achievements, for example, of 1-year-olds,

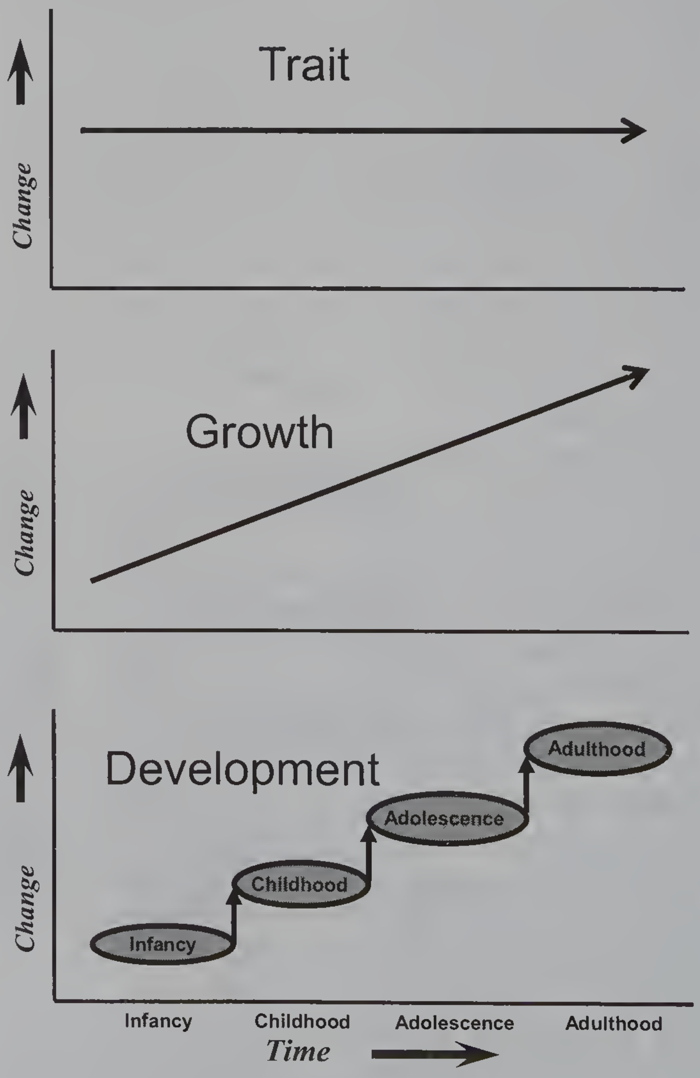


Figure 3. Personal change depicted as a trait, growth, or development.

2-year-olds, or 3-year-olds, similar to how intelligence quotient (IQ) tests are constructed. In contrast, the theoretical use of stage implies that there is a period of stability of functioning followed by a transition to a structurally different period of stability presumed to reflect more encompassing cognitive and social functioning. The classic examples of theoretical use of stages are in the writings of Freud and Piaget. Although there have been major revisions or rejections of these particular formulations, there are some generally accepted notions that within many domains individuals move from novices, to experts, to masters where they do not just do things better, they do things differently (Ericsson & Charness, 1994).

The general range of developmental changes has been extended well into adulthood and aging by the orientations of life span (Baltes, 1979) and life course theories (Elder, 1979) with their heavy emphasis on the importance of continuing alterations in the family, the workplace, and the historical epoch as individuals move into adulthood. The inability to separate individuals from context in the life-span models of adulthood provides a motivation to reconceptualize the importance of developmental context for younger individuals as well. The child or individual is not a unity and any model of the person also has to include the complex of psychological and underlying biological changes as well.

Contextual Model

Although developmental psychology is focused on individuals, it has become clear that understanding change requires an analysis of an individual's experience. Behavior, in general, and development, in particular, cannot be separated from the social context. Our understanding of experience has moved from a focus on primary caregivers to multiple other sources of socialization. There were many predecessors who felt that families, schools, neighborhoods, and culture had influences on development, but Bronfenbrenner turned these ideas into a comprehensive framework with predictions of how these settings affect the child but also how they affect each other. Although his terminology of microsystems, mesosystems, macrosystems, exosystems, and chronosystems may not be universally accepted, his principles that the family, school, and community are all intertwined in explaining any particular child's progress is now universally acknowledged (see Figure 4).

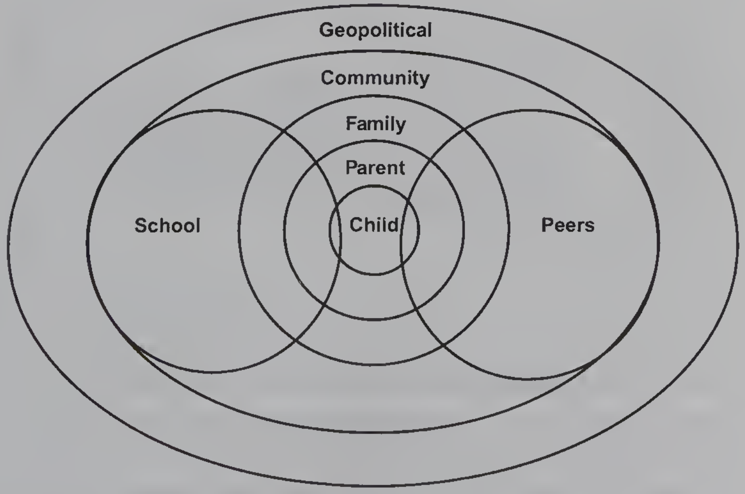


Figure 4. Social-ecological model of context.

Traditionally, social contacts were considered to expand from participation wholly in the family microsystem into later contact with the peer group and school system. Today, however, many infants are placed in out-of-home group child care in the first months of life. Each of these settings has its own system properties such that their contributions to the development of the child are only one of many institutional functions. For example, the administration of a school setting needs attention to financing, hiring, training of staff, and building maintenance before it can perform its putative function of caring for or educating children (Maxwell, 2009). Thus, a sociological analysis of such settings provides information about its ability to impact children.

Attention to the effects on children of changing settings over time must be augmented by attention to changing characteristics of individuals within a setting. Contemporary social models take a life course perspective that includes the interlinked life trajectories of not only the child but other family members (Elder, Johnson, & Crosnoe, 2003). For example, experience for the child may be quite different if the mother is in her teens with limited education, or in her 30s after completing professional training and entry into the job force.

Capturing the complex effects of multiple environmental situations has been a daunting enterprise requiring vast sample sizes to capture the unique contributions of each setting. An alternative methodology to dimensionalize the negative or positive quality of a child's experience has been the use of multiple or cumulative risk or promotive factor scores. For example, a set of data on the effects of a number of environmental variables on adolescent development was provided by a study of a large group of Philadelphia families (Furstenberg, Cook, Eccles, Elder, & Sameroff, 1999).

In the Philadelphia project 20 environmental factors were assessed and combined to approximate an ecological model containing six contextual subsystems. These were *Family Processes* that included support for autonomy, behavior control, parental involvement, and family climate; *Parent Characteristics* that included mental health, sense of efficacy, resourcefulness, and level of education; *Family Structure* that included the parents' marital status and socioeconomic indicators of household crowding and welfare status; *Family Management* composed of variables of institutional involvement, informal networks, social resources, and adjustments to economic pressure; *Peers* that included indicators of association with prosocial and antisocial peers; and *Community* that included census tract information on average income and educational level of the neighborhood, a parent report of neighborhood problems, and measures of the adolescent's school climate. In addition to the large number of ecological variables, we used a wide array of youth developmental outcomes in five domains: *Psychological Adjustment*, *Self-Competence*, *Conduct Problems*, *Extracurricular Involvement*, and *Academic Performance*.

For the environmental risk effects analyses each of the 20 variables was dichotomized with approximately one fourth of the families in the high-risk group and then the number of high risk conditions summed. When we examined the relation between the multiple risk factor score and the five adolescent outcomes, there were large declines in outcome with increasing risk and a substantial overlap in slope for each (Sameroff, 2006). Although this kind of epidemiological research does not unpack the processes by which each individual is impacted by contextual experience, it does document the multiple factors in the environment that are candidates for more specific analyses.

We also examined the effects of promotive influences in the Philadelphia study. Sameroff (1999) proposed that a better term for the positive end of the risk dimension would be *promotive* rather than protective factors. A promotive factor would have a positive effect in both high- and low-risk populations, which is far more common than a protective factor that only facilitates the development of high-risk children. We created a set of promotive factors by cutting each of our environmental variables at the top quartile, rather than the bottom, and summing them. The effects of the multiple promotive factor score mirrored the effects of the multiple risk score. Children from families with many promotive factors did substantially better than children from families with few promotive factors on each of our array of adolescent outcomes. For the youth in the Philadelphia sample, the more risk factors, the worse the outcomes, and the more promotive factors, the better the outcomes. In sum, context includes a constellation of environmental influences that have general effects on child development, fostering child development at one end and inhibiting it at the other.

Of great significance for the life course, these effects play out over time as a manifestation of the Matthew effect, "To the man who has, more will be given until he grows rich; the man who has not will lose what little he has" (Matthew 13:12). In a study of high- and low-IQ 4-year-olds we tracked their academic achievement through high school (Gutman, Sameroff, & Cole, 2003). The low-IQ group living in low contextual risk conditions consistently did better than the high-IQ group living in high risk conditions. Over time promotive or risky contextual effects either fostered or wiped out prior individual competence.

Regulation Model

The third component of the unified theory is the *regulation model* reflecting the systems orientation of modern science (Sameroff, 1983). The idea that that the child is in a dynamic rather than passive relationship with experience has become a basic tenet of contemporary developmental psychology. However, most of the rhetoric is about "self"-regulation. Whether it is Piaget's assimilation-accommodation model in cognition or Rothbart's (1981) reactivity and self-regulatory view of temperament, equilibration is primarily a characteristic native to the child. The context is necessary as a source of passive experiences that stimulate individual adaptation, but has no active role in shaping that adaptation. These views promote a belief that regulation is a property of the person. However, self-regulation mainly occurs in a social surround that is actively engaged in "other"-regulation. At the biological level the self-regulatory activity of genes is intimately connected to the other-regulatory activity of the surrounding cell cytoplasm. In Thelen's (1989) view of dynamic systems other-regulation is provided by the strange attractors of chaos theory. The self-regulation leading to an infant's neurologically based coordination of walking is constrained by the other-regulation of the child's muscle development, the strange attractor.

This issue of the developmental expansion of self-regulation to include other-regulation is captured by the *ice-cream-cone-in-a-can* model of development (Sameroff & Fiese, 2000) depicted in Figure 5. The developmental changes in the relation between individual and context are represented as an expanding cone within a cylinder. The balance between other-regulation and self-regulation shifts as the child is able to take on more and more responsibility for his or her own well-being. The infant, who at birth could not survive without the caregiving environment, eventually reaches adulthood and can become part of the other-regulation of a new infant, beginning the next generation.

It is parents who keep children warm, feed them, and cuddle them when they cry; peers who provide children with knowledge about the range and limits of their social behavior; and teachers who socialize children into group behavior as well as regulate cognition into socially constructed domains of knowledge. Although these other-regulators can be considered background to the emergence of inherent individual differences in regulatory capacities, there has been much evidence from longitudinal research among humans and cross-fostering studies in other animals that "self"-regulatory capacities are heavily influenced by the experience of regulation provided by caregivers. The capacity for self-regulation arises through the actions of others. This regulation by others provides the increasingly complex social, emotional, and cognitive experiences to which the child must self-regulate and the safety net when self-regulation fails. Children's cognition to a large extent is not derived from direct experiences with the environment but based on interpretations provided by others (Gelman, 2009). Moreover, these regulations are embedded not only in the relation between child and context but also in the additional relations between family and their cultural and economic situations (Raver, 2004). These regulatory systems range from the here-and-now experiences of parent–child interactions to governmental concern with the burden of national debt that will be passed on the next generation and to conservationists' concerns with the fate of the planet as a viable environment for future generations of humans.

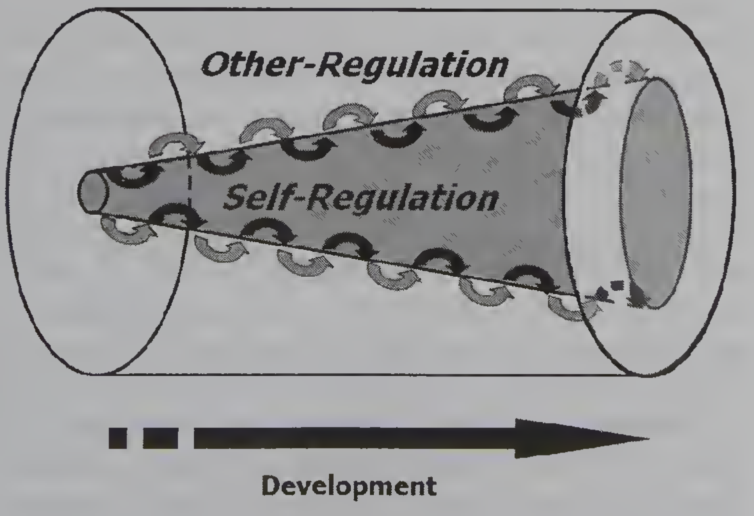


Figure 5. Transactional relations between self-regulation and other-regulation.

Early functional physiological self-regulation of sleep, crying, and attention are augmented by caregiving that provides children with regulatory experiences to help them quiet down on the one hand and become more attentive on the other. Sleep is an interesting example where biological regulation becomes psychological regulation through social regulation. As wakefulness begins to emerge as a distinct state it is expanded and contracted by interactions with caregivers who stimulate alertness and facilitate sleepiness. Although it remains an essential biological process, eventually it takes on a large degree of self-regulation as the child and then adult make active decisions about waking time and sleeping time. But this agentic decision making remains intimately connected with other-regulation in terms of the demands of school and work for specific periods of wakefulness.

Robert Emde and I with a group of colleagues (Sameroff & Emde, 1989) in an attempt to describe mental health diagnoses for infants argued for a position that infant diagnoses could not be separated from relationship diagnoses. Our point was that in early development life is a "we-ness" rather than an "I-ness." The developmental and clinical question in this case is when does diagnosis become individualized, at what stage does a child have a self-regulation problem instead of an other-regulation problem? One answer is to identify the point in development when areas of self-regulation become independent of initial regulatory contexts and are carried into new relationships. Children who have imaginary playmates provide an interesting perspective on the relation between self- and other-regulation. The more preschoolers engaged in fantasy and pretense, the more sophisticated their theory of mind (Taylor & Carlson, 2009).

Generally, research into self-regulation has focused on part processes, such as emotion or attention. Such empirical isolation obscures the larger picture in which many interacting systems are playing significant roles. Without regulation provided by the social context, for example, nutrition and temperature, the young child would not survive to engage in emotional or attentional processes.

Transactional Regulation

The previous discussion of the need for a construct of other-regulation to complete an understanding of self-regulation leads now to how the relation between self and other operates developmentally and for this we turn to the transactional model (Sameroff & Chandler, 1975). Transactions are omnipresent. Everything in the universe is affecting something else or is being affected by something else. In the transactional model the development of the child is a product of the continuous dynamic interactions of the child and the experience provided by his or her social settings. What is core to the transactional model is the analytic emphasis placed on the interdependent effects of the child and environment and is depicted in the bidirectional arrows between self and other in Figure 5.

In a recent book on the topic (Sameroff, 2009), a number of researchers documented transactional processes in cognitive and social-emotional domains where agents in the family, school, and cultural contexts altered the course of children's development in both positive and negative directions. Transactional examples have been typically in the behavioral domain with an emphasis on parent–child mutual exacerbations producing problem behavior in both partners (Patterson, 1986). More recently, transactions have been recognized in teacher–student relationships where the effects of the teacher on the child in one grade will change the reaction of the teacher in the next moving the student to higher or lower levels of competence (Morrison & Connor, 2009). Multilevel transactions have also been documented where not only the parent and child are transacting with each other but both are also transacting with cultural practices (Bornstein, 2009).

Vygotsky's (1978) zone of proximal development is analogous to transactional other-regulation in cognitive development. Successful socialization and particularly good education is based on fitting experience to the developmental status of the child. As children create their understanding of the world, the world is made more complex through steps in a curriculum to move them along toward some societal goal of mature thought. Arithmetic is an excellent example where as soon as children learn to add, they are required to learn to subtract, following which they are taught to multiply and divide. Each step is a transactional regulation of the environment by the teacher to keep one step ahead of the child's mathematical regulation. Similarly, in the social realm increases in social responsibility are paced to the success of the adjustment to previous levels of responsibility (Rogoff, 2003).

In a more popular vein Gladwell (2008) describes the life course of a number of eminent individuals in sports, commerce, and technology, where equally competent children did not achieve similar greatness because of the lack of social, educational, or technological possibilities. In each case initial advantage scaffolded the child to be able to elicit and make use of a series of opportunities documenting the transactional progression that eventually led to eminence.

Representational Model

Representations are encodings of experience. They are a more or less elaborated internal summary of the external world. They include the cognitive representations where the external world is internalized, the social representations where relationships become working models, the cultural representations of different ethnicities or social classes, and the developmental theories discussed here. Representations are obviously not the same as what they represent. They have an adaptive function of bringing order to a variable world, producing a set of expectations of how things should fit together.

We have long been familiar with such representations as perceptual constancy in which objects are perceived as being a certain size even when the sensory size is manipulated. In such a summation certain aspects are selected and others ignored. In the representation of a square for example, the size, color, and texture of the square object may be ignored. Analogously, when representations are made of a social object such as a parent, certain features are included in the representation and others are ignored. Research using the adult attachment interview (Main & Goldwyn, 1984) has found that representations of parents are often idealized, where only positive aspects are included in the mental model. Although the links between the quality of representations of child–parent relationships during infancy and those during adulthood are far from direct, early working models of attachment do seem to have long-term consequences for adult development (Sroufe, Egeland, Carlson, & Collins, 2005).

Similarly, parents create representations of their children that emphasize certain aspects, deemphasize others, and have stability over time independent of the child's actual characteristics. We had parents rate their infants' temperament during the 1st year of life following a structured interaction sequence (Seifer, Sameroff, Barrett, & Krafchuk, 1994). We also had them rate the temperament of six unfamiliar infants engaged in the same interaction sequence. The average correlation in temperament ratings of the unfamiliar infants between mothers and trained observers was .84 with none below .60. The average correlation in temperament ratings between mothers and trained observers for their own children was .35 with a range down to −.40. Mothers were very good raters of other people's children but very poor raters of their own due to the personal representations that they imposed on their observations. Documenting such differences in parent representations would be of no more than intellectual interest, if there were not consequences for the later development of the child. For example, infants whose mothers perceived them as problematic criers during infancy increased their crying during toddlerhood and had higher problem behavior scores when they were preschoolers (McKenzie & McDonough, 2009).

Individual well-being is also a result of meaningful cultural engagement with desirable everyday routines that have a script, goals, and values (Weisner, 2002). Meaningfulness, a key component of cultural analyses, is primarily found in coherent representations. Evidence of a positive effect of meaning systems can be found in Fiese and Winter's (2009) descriptions of how family routines provide a narrative representation for the rest of the family members that allows the whole to continue adaptive functioning despite the variability in the behavior of the parts. Evidence of a negative effect of lack of meaningfulness is in a study of native Canadian youth who showed much higher levels of suicide and other problem behavior when there were large inconsistencies in cultural continuity from one generation to another (Chandler, Lalonde, Sokol, & Hallett, 2003). The order or disorder in a family or society's representation of itself affects the adaptive functioning of its members.

Unifying the Theory of Development

Now that the four models necessary for a theory of development have been described, I can proceed to integrate them into a comprehensive view that contains most known influences on life trajectories. I will begin with a structural depiction of the components of the personal and contextual models containing all the pieces relevant to development. I will then add the regulation and change component of the personal model to capture the processes that produce the life course and then finish the unified theory with an overlay of the representational model.

Structural Formulation

The self is composed of a set of interacting psychological and biological processes. The psychological domains overlap in cognitive and emotional realms of intelligence, mental health, social competence, and identity, among others. These are depicted as the set of grey, overlapping circles comprising the psychological part of the self in Figure 6. Each of these psychological domains is subserved by and interacts with a set of interacting biological processes, including neurophysiology, neuroendocrinology, proteomics, epigenomics, and genomics that are depicted as a set of black, overlapping circles. Together the gray and black circles comprise the *biopsychological* self system. This self-regulation system interacts with the other-regulation system, depicted by the surrounding white circles, representing the many interacting settings of the social ecology, including family, school, neighborhood, community, and overarching geopolitical influences. Taken together the three sets of overlapping circles comprise the *biopsychosocial* aspects of the individual in context.

Process Formulation

The process formulation adds the personal change time dimension to the biopsychosocial model, which can be viewed as either a *growth model*, where the biopsychological aspects increase quantitatively over time but there is no change in their interrelationships as in the cone image (see Figure 5) or a *developmental model*, where the aspects have qualitative shifts in organization in which there are changing relations among the biopsychosocial aspects (see Figure 7).

Evolutionary theory has provided a fruitful analog for understanding the transitions that lead from one developmental stage to another. As opposed to the gradualist understanding of evolutionary changes originally proposed by Darwin that would look like the growth model, Eldredge and Gould (1972) argued that evolution was characterized by continuity evidenced in long periods of stasis where there were only modest changes alternating with discontinuity where there were short periods of rapid change that they labeled *punctuated equilibrium*. The implication was that there was a balance

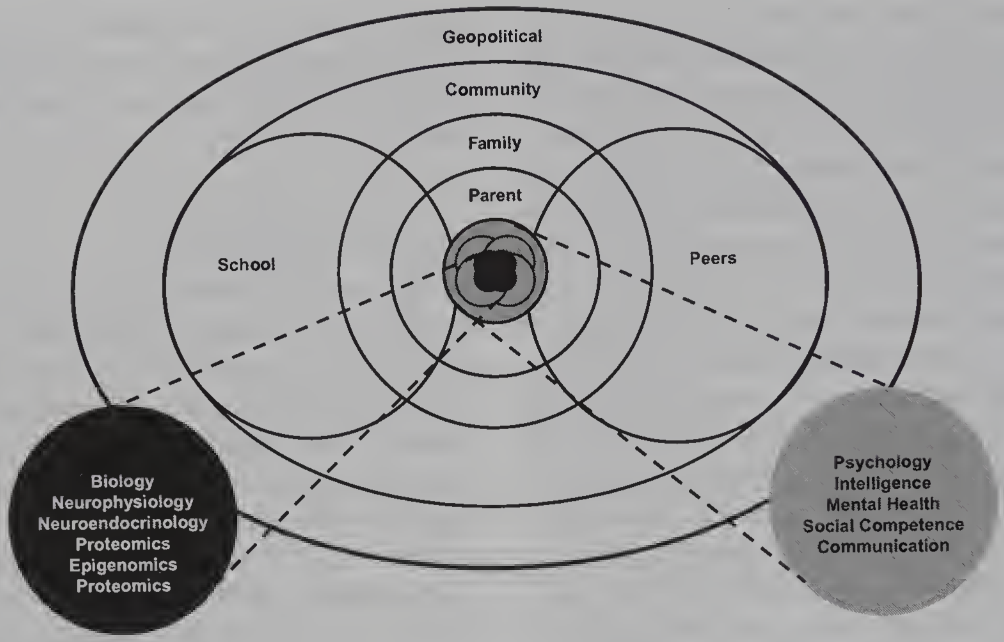

Geopolitical
Community
Family
Parent
School
Peers
Biology
Neurophysiology
Neuroendocrinology
Proteomics
Epigenomics
Proteomics
Psychology
Intelligence
Mental Health
Social Competence
Communication

Figure 6. Biopsychosocial ecological system.

between species and their ecosystems until it was interrupted by either large changes in the species or large changes in the environment that required a new equilibration. In terms of understanding developmental discontinuities in the individual, we would need to search for such changes in the child or the context that create pressures for a new equilibration. These forces are represented by the up and down arrows around points of inflection in Figure 7.

One of the most commonly accepted transitions has been the 5- to 7-year shift in cognition originally documented in 21 behavioral domains by White (1965) and accentuated in the work of Piaget. Thirty years later Sameroff and Haith (1996) and a group of contributors reexamined this transition but also asked if there were contextual changes during this age period. We reached the conclusion that there was a 5- to 7-year shift in the child if by 5 we meant 3 and by 7 we meant 10. This answer reflects the study of what might be called "part processes." If one asks whether 5-year-olds can attend, remember, have emotions, engage in social interactions, and even take charge of social interactions, the answer is yes. If one asks whether 5-year-olds can fully integrate their physical, cognitive, emotional, and social worlds, the answer is no. But neither can 7-year-olds. So what is the punctuation between the ages of 5 and 7? On average 5- to 7-year-olds can integrate several behaviors that permit the beginnings of formal education in most cultures in the world—increased cognitive ability, the ability to sit still, and the ability to pay attention. Some children have these capacities much earlier, but the requirements for successful participation in the school setting require all three plus a number of others. White's (1996) more recent conclusion was that, "what happens to children between 5 and 7 is not the acquisition of an absolute ability to reason; it is an ability to reason with others and to look reasonable in the context of society's demands on the growing child to be cooperative and responsible (p. 27)." In Figure 7 there are up arrows from self to other reflecting child advances, but there may be more powerful influences from other to self where society does the developmental punctuation by requiring the child to spend most of the day in school rather than at home. From this perspective the stages of infancy, childhood, adolescence, and adulthood could be

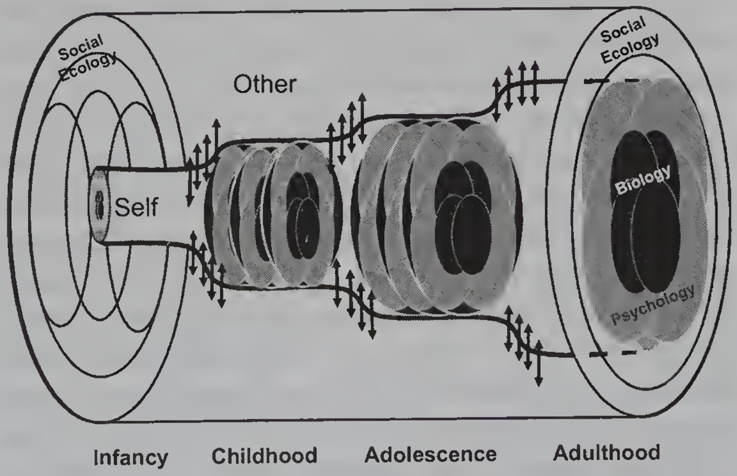

Social Ecology
Other
Self
Social Ecology
Biology
Psychology
Infancy
Childhood
Adolescence
Adulthood

Figure 7. Unified theory of development including the personal change, context, and regulation models.

relabeled the home stage, the elementary school stage, the secondary school stage, and the work and new family stage.

Similar analyses can be applied to the punctuations that occur in the transition to adolescence or adulthood. It is the relation between shifts in the child and shifts in the context that mark new stages. Puberty is a biological achievement of the child but adolescence is a socially designated phase between childhood and adulthood (Worthman, 1993). Puberty is universal but adolescence is not, either in historical or cross-cultural perspective. In many cultures adolescence is directly tied to biological changes but in modernizing cultures it is more closely tied to age-based transitions into middle and high schools. Depending on the culture sexual participation can be encouraged at an early age before biological maturity or discouraged until individuals are well into adulthood. These pressures from changes in the child and the context are represented by the up and down arrows around the adolescent transition in Figure 7. In western societies, adolescence is generally recognized but the quality of the adolescent experience is quite variable and may be heavily dependent on stage–environment fit. Depending on the particular family or school system, desires for autonomy and intimacy can be fostered or thwarted moving the adolescent into better or worse future functioning. Negative psychological changes associated with adolescent development often result from a mismatch between the needs of developing adolescents and the opportunities afforded them by their social environments (Eccles et al., 1993).

The unified theory depicted in Figure 7 combines the personal change, contextual, and regulation model, but it would become overly complex to add the representational model to the figure, as well. Suffice it to say that representation suffuses every aspect of the model in the interacting identities, attitudes, beliefs, and attributions of the child, the family, the culture, and the organizational structure of social institutions. Moreover, the way developmental science conceptualizes the child may be only one of a number of possible cultural inventions (Kessen, 1979). The most important representation for current purposes is captured in the depiction of a unified theory of development. Like most theories the unified view does not make specific predictions but does specify what will be necessary for explaining any developmental phenomena. It is a reversal of the usual bottom-up empirical stance where the researcher maintains as narrow focus as possible unless forced to enlarge the scope by some contradictory findings. The top-down theoretical stance is that researchers need to be aware that they are examining only a part of a larger whole consisting of multiple interacting dynamic systems.

Future of Nature Versus Nurture

Current Nature Ascendance

The current ascendance of research using new biological measures of individual differences is the result of the interdisciplinary collaboration that Parke (2004) had indicated was essential to the advance of developmental research. These advances in molecular genetics, endocrinology, and neurology are being rapidly integrated into psychological research. The good news is that the new science is no longer based on the reductionist models of the past where linear progressions were proposed between biological entities such as genes or neurotransmitters and psychological function. In each domain multidirectional models are replacing unidirectional ones with a growing emphasis on gene–environment interactions, epigenome–experience transactions, and brain plasticity. These advances are relationship based, requiring increasingly complicated systems analyses to capture the multiple part–whole processes underlying developmental change. Nurture, for example, the environment of the gene, the environment of the cell, and the environment of the organism, are incorporated into advanced analyses of the contribution of context at every level of analysis. It is striking that the nonreductionist systems thinking that those who define psychology as a natural science have avoided is a now a central part of their colleague disciplines of biology and physics. Developmental science is benefiting from advances in the natural sciences at the theoretical as well as the empirical level.

Next Resurgence of Nurture

A renewed emphasis on the importance of nurture is underway. Again, it is a dialectical result of the inability of appeals to human nature to explain fully developmental pathways. There remain large amounts of unexplained variance. The nurture resurgence is implicit in the new directions for biological sciences such as epigenomics, described above, and will become explicit with a more powerful appreciation of the perspectives on human development provided by social sciences beyond psychology. The core element in each interdisciplinary effort is that successful developmental predictions from psychological measures are highly contingent on the social or biological context. Two of the major ingredients needing integration into a unified developmental science are the *opportunity structure* construct from sociology and economics and the *meaning making* construct from anthropology.

The important perspective that sociology adds to developmental science is that individuals are embedded in networks of relationships that constrain or encourage different aspects of individual behavior. Social institutions like families, schools, and the workplace are composed of roles that children come to understand and fill. In this view individual differences, the core of psychological concern, are limited by role demands in predicting developmental outcomes. Economists are interested in what keeps economies going and individual behavior is viewed through the lens of financial choices. The part of economics most relevant to behavioral development is the availability of an opportunity structure. Once again the predictive power of individual differences is constrained by the availability of such resources as educational systems, job choices, and social mobility that determines whether individuals have the option to use their prior competencies or not. Anthropology is indeed interested in cultural differences in behavior, but equally important for understanding development are differences in meaning systems, that is, how different cultures think about their practices. The same behavior can have quite different meanings and quite different behaviors can have the same meaning in different cultures. Again the predictive power of individual differences is constrained by how different cultures value and proscribe different behaviors.

Development of the Developmentalist

I began this article proposing that the study of the development of our field would illuminate our study of the development of individuals. Up until the 1960s *child psychologist* was the predominant label for researchers with children and the main focus was on identifying measures of stable intelligence and personality traits that would be predictive of adult performance. In the 1960s and 1970s we became *developmental psychologists* as organizational principles and emergents dominated the rhetoric around the cognitive revolution and attachment theory. During the 1980s and 1990s we reframed ourselves as *developmental scientists* when we gained a fuller appreciation of the contribution of biology and the social ecology to psychological growth. In the new millennium we again are changing our self description to *developmental systems theorists* as multilevel biopsychosocial dynamic systems are becoming the framework for understanding human change over time and statisticians are providing tools that are closer approximations to the complexity of our data.

With regard to what we have learned about nature and nurture, the future challenge is not to find new arguments for one or the other but to create a developmental model where advances in the study of both individual and context are expected and hoped for. I have proposed such a biopsychosocial unified theory of development that I hope will be useful for future research in human development. Over time the body changes, the brain changes, the mind changes, and the environment changes along courses that may be somewhat independent of each other and somewhat a consequence of experience with each other. It should be a very exciting enterprise to fill in the details of how biological, psychological, and social experiences foster and transform each other to explain both adaptive and maladaptive functioning across the life course.

Coming full circle to the dialectical principles of the yin–yang model, there are continuities as scientists concerned with greater differentiations within our biological and social experience continue to push our understanding of both nature and nurture. But there are increasing discontinuities with the rhetoric of the past as many more developmentalists realize that neither nature nor nurture will provide ultimate truths and neither can be an end in itself. Instead, each can explain the influences of the other because in the end neither can exist without the other. They mutually constitute each other through their unity and interpenetration of opposites. The schematic depiction of the unified theory of development provides an integrated way of looking at things, but also for things. Although we all have a strong desire for straightforward explanations of life, development is complicated and models for explaining it need to be complicated enough to usefully inform our understanding.

> Everything should be as simple as possible, but not simpler.
>
> Albert Einstein

References

Baltes, P. B. (1979). Life-span developmental psychology: Some converging observations on history and theory. In P. B. Baltes & O. G. Brim Jr. (Eds.), *Life-span development and behavior* (Vol. 2, pp. 256–281). New York: Academic Press.

Bornstein, M. H. (2009). Toward a model of culture-parent-child transactions. In A. Sameroff (Ed.), *The transactional model of development: How children and contexts shape each other* (pp. 139–161). Washington, DC: American Psychological Association.

Boulding, K. (1956). General systems theory—The skeleton of science. *Management Science*, *2*, 197–208.

Breland, K., & Breland, M. (1961). The misbehavior of organisms. *American Psychologist*, *16*, 681–684.

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist*, *32*, 513–531.

Bronfenbrenner, U., Kessel, F., Kessen, W., & White, S. (1986). Toward a critical social history of developmental psychology: A propaedeutic discussion. *American Psychologist*, *41*, 1218–1230.

Chandler, M. J., Lalonde, C. E., Sokol, B. W., & Hallett, D. (2003). Personal persistence, identity development, and suicide: A study of native and non-native North American adolescents. *Monographs of the Society for Research in Child Development*, *68*(2, Serial No. 273).

Cicchetti, D., & Rogosch, F. A. (1996). Equifinality and multifinality in developmental psychopathology. *Development and Psychopathology*, *8*, 597–600.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cole, M. (2006). Culture and cognitive development in phylogenetic, historical, and ontogenetic perspective. In D. Kuhn & R. Siegler (Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception and language* (6th ed., pp. 636–683). New York: Wiley.

Defries, G. E., & McLearn, J. C. (1973). *Introduction to behavioral genetics*. San Francisco: W.H. Freeman.

Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., et al. (1993). Development during adolescence: The impact of stage-environment fit on young adolescents' experiences ins schools and in families. *American Psychologist*, *48*, 90–101.

Elder, G. H., Jr. (1979). Historical change in life patterns and personality. In P. Baltes & O. Brim Jr. (Eds.), *Life-span development and behavior* (pp. 117–159). New York: Academic Press.

Elder, G. H., Jr., Johnson, M. K., & Crosnoe, R. (2003). The emergence and development of life course theory. In J. T. Mortimer & M. J. Shanahan (Eds.), *Handbook of the life course* (pp. 3–19). New York: Kluwer Academic/Plenum.

Eldredge, N., & Gould, S. J. (1972). Punctuated equilibria: an alternative to phyletic gradualism. In T. J. M. Schopf (Ed.), *Models in paleobiology* (pp. 82–115). San Francisco: Freeman Cooper.

Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist*, *49*(8), 725–747.

Erikson, E. H. (1959). Identity and the life cycle. *Psychological Issues*, *1*, 1–171.

Fiese, B. H., & Winter, M. A. (2009). The dynamics of family chaos and its relation to children's socio-emotional well being. In G. W. Evans & T. D. Wachs (Eds.), *Chaos and its influence on children's development: An ecological perspective* (pp. 55–76). Washington, DC: American Psychological Association.

Furstenberg, F. F., Jr., Cook, T. D., Eccles, J., Elder, G. H., Jr., & Sameroff, A. (1999). *Managing to make it: Urban families and adolescent success.* Chicago: University of Chicago Press.

Galton, F. (1876). The history of twins, as a criterion for the relative powers of nature and nurture. *Journal of the Royal Anthropological Institute*, *5*, 391–406.

Gelman, S. A. (2009). Learning from others: Children's construction of concepts. *Annual Review of Psychology*, *60*, 115–140.

Gladwell, M. (2008). *Outliers: The story of success.* New York: Little, Brown.

Gottlieb, G. (1992). *Individual development and evolution: The genesis of novel behavior*. New York: Oxford University Press.

Gutman, L. M., Sameroff, A. J., & Cole, R. (2003). Academic growth curve trajectories from first to twelfth grades: Effects of multiple social risk and preschool child factors. *Developmental Psychology*, *39*, 777–790.

Hinde, R. A. (1997). *Relationships: A dialectical perspective.* Hove, UK: Psychology Press.

Huston, A. C. (2008). From research to policy and back. *Child Development*, *79*, 1–12.

Kessen, W. (1962). "Stage" and "structure" in the study of children. In W. Kessen & C. Kuhlman (Eds.), Thought in the young child: Report of a conference with particular attention to the work of Jean Piaget. *Monographs of the Society for Research in Child Development*, *27*(2, Serial No. 83).

Kessen, W. (1979). The American child and other cultural inventions. *American Psychologist*, *34*, 815–820.

Lorenz, K. (1950). Innate behaviour patterns. In *Symposia of the society for experimental biology no. 4: Physiological mechanisms in animal behaviour*. New York: Academic Press.

Main, M., & Goldwyn, R. (1984). Predicting rejection of their infant from mother's representation of her own experience: Implications for the abused and abusing intergenerational cycle. *Child Abuse and Neglect*, *8*, 203–217.

Maxwell, L. E. (2009). Chaos outside the home: The school environment. In G. W. Evans & T. D. Wachs (Eds.), *Chaos and its influence on children's development: An ecological perspective* (pp. 117–136). Washington, DC: American Psychological Association.

McKenzie, M. J., & McDonough, S. C. (2009). Transactions between perception and reality: Maternal beliefs and infant regulatory behavior. In A. Sameroff (Ed.), *The transactional model of development: How children and contexts shape each other* (pp. 35–54). Washington, DC: American Psychological Association.

Morrison, F. J., & Connor, C. M. (2009). The transition to school: Child instruction transactions in learning to read. In A. Sameroff (Ed.), *The transactional model of development: How children and contexts shape each other* (pp. 183–201). Washington, DC: American Psychological Association.

Parke, R. D. (2004). The Society for Research in Child Development at 70: Progress and promise. *Child Development, 75*, 1–24.

Patterson, G. R. (1986). Performance models for antisocial boys. *American Psychologist, 41*, 432–444.

Princeton Review. (2009). *Cracking the GRE.* New York: Princeton Review Books.

Raver, C. C. (2004). Placing emotional self-regulation in sociocultural and socioeconomic contexts. *Child Development, 75*, 346–353.

Riegel, K. F. (1976). The dialectics of human development. *American Psychologist, 31*, 689–700.

Rogoff, B. (2003). *The cultural nature of human development.* Oxford, UK: Oxford University Press.

Rothbart, M. K. (1981). Measurement of temperament in infancy. *Child Development, 52*, 569–578.

Sameroff, A. J. (1983). Developmental systems: Contexts and evolution. In W. Kessen (Series Ed.) & P. H. Mussen (Vol Ed.), *Handbook of child psychology: Vol. 1. History, theories, and methods.* (pp. 238–294). New York: Wiley.

Sameroff, A. J. (1999). Ecological perspectives on developmental risk. In J. D. Osofsky & H. E. Fitzgerald (Eds.), *WAIMH handbook of infant mental health: Vol. 4. Infant mental health groups at risk* (pp. 223–248). New York: Wiley.

Sameroff, A. J. (2006). Identifying risk and protective factors for healthy youth development. In A. Clarke-Stewart & J. Dunn (Eds.), *Families count: Effects on child and adolescent development* (pp. 53–76). Cambridge, UK: Cambridge University Press.

Sameroff, A. J. (Ed.) (2009). *The transactional model of development: How children and contexts shape each other.* Washington, DC: American Psychological Association.

Sameroff, A. J., & Chandler, M. J. (1975). Reproductive risk and the continuum of caretaking casualty. In F. D. Horowitz, M. Hetherington, S. Scarr-Salapatek, & G. Siegel (Eds.), *Review of child development research* (Vol. 4, pp. 187–244). Chicago: University of Chicago Press.

Sameroff, A. J., & Emde, R. N. (Eds.) (1989). *Relationship disturbances in early childhood: A developmental approach.* New York: Basic Books.

Sameroff, A. J., & Fiese, B. H. (2000). Transactional regulation: The developmental ecology of early intervention. In J. P. Shonkoff & S. J. Meisels (Eds.), *Early intervention: A handbook of theory, practice, and analysis* (2nd ed., pp. 135–159). New York: Cambridge University Press.

Sameroff, A. J., & Haith, M. (Eds.). (1996). *The five to seven year shift: The age of reason and responsibility.* Chicago: University of Chicago Press.

Seifer, R., Sameroff, A. J., Barrett, L. C., & Krafchuk, E. (1994). Infant temperament measured by multiple observations and mother report. *Child Development, 65*, 1478–1490.

Seligman, M. E. P. (1970). On the generality of the laws of learning. *Psychological Review, 77*, 406–418.

Sroufe, L. A., Egeland, B., Carlson, E. A., & Collins, W. A. (2005). *The development of the person: The Minnesota study of risk and adaptation from birth to adulthood.* New York: Guilford.

Taylor, M., & Carlson, S. M. (2009). The relation between individual differences in fantasy and theory of mind. *Child Development, 68*, 436–455.

Thelen, E. (1989). Self-organization in developmental processes: Can systems approaches work? In M. R. Gunnar & E. Thelen (Eds.), *Systems and development: The Minnesota symposia on child psychology* (Vol. 22, pp. 77–117). Hillsdale, NJ: Erlbaum.

von Bertalanffy, L. (1968). *General system theory* (Rev. ed.). New York: Braziller.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Watson, J. D. (1914). *Behavior: An introduction to comparative psychology.* New York: Holt.

Weisner, T. S. (2002). Ecocultural understanding of children's developmental pathways. *Human Development, 45*, 275–281.

Werner, H. (1948). *Comparative psychology of mental development.* New York: International University Press.

Werner, H. (1957). The concept of development from a comparative and organismic point of view. In D. B. Harris (Ed.), *The concept of development* (pp. 125–148). Minneapolis: University of Minnesota Press.

White, S. H. (1965). Evidence for a hierarchical arrangement of learning processes. *Advances in Child Behavior and Development, 2*, 187–220.

White, S. H. (1996). The child's entry into the "Age of Reason." In A. J. Sameroff & M. M. Haith (Eds.), *The five to seven year shift: The age of reason and responsibility* (pp. 17–30). Chicago: University of Chicago Press.

Worthman, C. M. (1993). Biocultural interactions in human development. In M. E. Pereira & L. A. Banks (Eds.), *Juvenile primates: Life history, development, and behavior* (pp. 339–358). New York: Oxford University Press.

Child Development, January/February 2010, Volume 81, Number 1, Pages 23–27

Introduction to the Special Section on The Effects of Early Experience on Development

Nathan A. Fox
University of Maryland

Michael Rutter
Institute of Psychiatry, Kings College London

One cardinal principle in developmental psychology is that early experience has a profound effect upon human development. Infants, once viewed as passive recipients of stimulation, are now understood to be active participants in the physical and social world that surrounds them. Learning takes place from a very early age and sets the course for trajectories of either adaptive or maladaptive behavior. A corollary to this principle is that there are certain periods during early development when experiences have a more significant effect than others. These periods, called sensitive or critical periods, are thought of as windows of opportunity during which certain types of experience have a foundational effect upon the development of skills or competencies. This special section of *Child Development* takes advantage of major advances in neuroscience, genetics, and improved developmental and statistical methods for studying the effects of early experience to provide readers with a broad range of review and empirical studies on this general topic. The special section is divided into seven parts: the first comprises two invited articles (Fox, Levitt, & Nelson, 2010; Meaney, 2010) that lay out important advances in neuroscience and molecular genetics. The second part provides five articles on fetal development and early experience. The third part contains three articles on perinatal experiences including low birth weight and long-term consequences on health and development, while the fourth part contains four articles on the effects of severe psychosocial deprivation, neglect, and abuse. The fifth part provides the reader with two articles on early experience and stress reactivity, while the sixth part presents three empirical articles examining basic mechanisms linking early experiences including poverty and adoption to behavior. Finally, Jack Shonkoff provides a commentary on the section with an eye toward thinking about the policy implications of the science.

The importance of early experience and the identification of sensitive or critical periods have a long history within developmental research. Much of the early work was on the effects of early handling and stress reactivity by developmental psychobiologists such as Seymour Levine (Levine, 1957) and Victor Denenberg (Denenberg, 1964). Their work, primarily in the rodent, illustrated the effects of early handling on later learning and reactivity. The notion of critical periods was introduced to developmental psychology by ethologists, such as Hess and Lorenz (e.g., Hess, 1964) with work on imprinting. Thinking about the effects of early experience on development has been solidified with advances in neuroscience that have described the pattern of brain development during the early months and years of life and the role that experience has in shaping development. Two areas of research have changed the way we conceptualize the effects of early experience and the influences of biology and nurture. The first is the work in neuroscience, done almost exclusively in rodents and nonhuman primates, on the effects of early experience on brain development. The second area is the revolution in our thinking about the role of experience in the action of genes and molecular genetics.

Neuroscience has provided an important foundation for our thinking about the role of early experience. Two lines of investigation are notable here. First, the descriptive work of Huttenlocher (Huttenlocher & Dabholkar, 1997) and Rakic (Granger, Tekaia, Le Sourd, Rakic, & Bourgeois,

---

Correspondence concerning this article should be addressed to Nathan A. Fox, Department of Human Development, University of Maryland, College Park, MD 20742. Electronic mail may be sent to fox@umd.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8101-0003

1995) illustrated changes in synaptic density, increases in neural connections, and the subsequent pruning or decreases in synaptic number that occurred during the postnatal period and differed in timing of occurrence as a function of brain region. Sensory and perceptual regions displayed these changes early in life while these changes occurred later in areas of the brain involved in higher cognition. Neuroscientists such as Greenough (Greenough, Black, & Wallace, 1987) argued that the pattern of synaptic ''blooming'' and ''pruning'' was in part a function of experience and the quality of stimulation that the organism was exposed to during specific time periods. A second and related line of work in the area of perceptual development by Hubel and Wiesel (1970) described in specific detail how certain types of sensory/perceptual experience were necessary, at particular early periods of postnatal development, for the mature functioning of the visual cortex. The animal had to engage actively in the perceptual world to stimulate the development of neural structures that underlay typical depth perception and other perceptual abilities. Suffice it to say, these studies and others supported what developmental psychologists thought for many years—that the quality and timing of early experience was critical for typical brain and behavioral development. The article by Fox et al. (2010) in this special issue provides the reader with an overview of these issues and a foundation for understanding the effects of both timing and quality of experience on psychological development.

Since the discovery and description of DNA, the importance of genes for understanding human behavior has been a continued subject of debate. For many years, this debate was informed by twin and adoption studies and the field of behavioral genetics. Data from these studies provided estimates of heritability of specific human personality traits or cognitive processes and at the same time, attempted to model the effects of shared and nonshared experience on the developing child. This work has been useful in maintaining a voice in the debate as to whether development was more influenced by ''nature'' versus nurture. Two recent approaches, however, illustrate that framing the question about early experience in this manner may no longer be valid. The first approach is one that models the different ways in which gene and environment can interact (Rutter, 2007). Among these possibilities is the interaction between individuals with a particular genotype and experience. Individuals carrying a particular allelic structure (e.g., if they are homozygous for the short allele of the 5HTT gene) are more susceptible to certain types of experiences and when exposed to these events have trajectories of development that are different from individuals with similar genetic makeup but different experience. A classic study by Caspi et al. (2003) illustrated this gene by environment interaction within the realm of psychiatric disorders. Individuals who were homozygous for the short allele of 5HTT and who experienced heightened life stress were more likely to have a diagnosis of depression compared to individuals with the same genetic makeup but who experienced less life stress. The obvious implications are that both gene AND environment are playing critical synergistic roles in understanding developmental outcome.

A second, more recent approach is one driven by advances in molecular biology and encompasses the field of epigenetics. Research in this area, initially the result of work on the genetics of cancer, argues that experience influences the cellular machinery of the gene, changing in some instances, the expression of the gene and the genome itself. It is the essence of a gene by environment interaction. Experience modifies the gene and its actions creating changes in behavior and in some instances being transmitted down through genetic action to subsequent generations. This work has been brought to the field of developmental psychology by Michael Meaney and his colleagues (Meaney, 2001). Their research has illustrated at the behavioral and genetic levels how variations in caregiving experiences affect stress reactivity and learning in the rodent. Meaney's article in this special section (Meaney, 2010) is an attempt to provide the readers of *Child Development* with a basic introduction to the field of epigenetics and to the power and potential of this approach to studying the effects of early experience on development.

Research into the development of the central nervous system (CNS) has focused, as well, on the period between conception and birth. By the time a full-term infant is born the basic wiring of the CNS has been completed, the neurons that form the different layers of the CNS have all reached their destinations, and a good deal of the basic wiring in the brain has been achieved. Whereas much of the ontogeny of CNS during fetal development is under genetic control, these genes and their actions are affected by the fetal environment. At one time, this environment was thought to be protected by an impervious placenta, but we now know that exposure to a range of substances can have neurotoxic effects and that maternal psychological state as well affects fetal development. Although much of developmental psychology's efforts have been toward describing learning and development and the effects of experience on cognitive and social development postbirth, the quality of the fetal environment and the timing of exposure during fetal development to a wide range of substances appear to play a significant role in typical or atypical postnatal and human development. Five articles are included in the special section illustrating the importance of the fetal environment on subsequent development. DiPietro et al. (2010) provide the reader with a roadmap to how one studies typical fetal development across pregnancy and the meaning of variability in fetal responses for newborn CNS maturation. Two articles (Price, Grosser, Plomin, & Jaffee, 2010; D'Onofrio et al., 2010) provide readers with an example of the consequences of fetal exposure to maternal smoking. D'Onofrio et al. (2010) examine the long-term consequences on academic achievement for children whose mothers smoked during pregnancy, with the important finding that the supposed prenatal environmentally mediated effects are probably largely a function of a shared genetic liability. Price et al. (2010) demonstrate an important gene by environment mechanism for understanding the effects on intrauterine growth, and D'Onofrio et al. (2010) examine the long-term consequences on academic achievement for children whose mothers smoked during pregnancy, with the important finding that the supposed prenatal environmentally mediated effects are probably largely a function of a shared genetic liability. Davis and Sandman (2010), using data from an ongoing longitudinal study, found that the timing of cortisol levels during gestation were associated with differences in the babies' developmental level at 12 months. Finally, Hay, Pawlby, Waters, Perra, and Sharp (2010) examine the effects of maternal antenatal depression on children's antisocial behavior. Remarkably, maternal depression during pregnancy above other risk factors during pregnancy and after birth significantly predicted child psychiatric problems.

Basic research into brain development in the human fetus prior to term delivery has also informed thinking about the role of early experience amongst infants born early. Neonatology has developed the methods to increase the survival of very low birth weight infants who are, in most cases, also very premature. Although mortality has decreased in this population, morbidity due to prenatal or neonatal complications is a major issue for this population of children. But even among those very low birth weight infants who do not undergo significant medical complications, the nonfetal environment into which they are born and the stimulation of the neonatal intensive care unit are clearly nonintended with regard to typical biological growth of the CNS. Two articles in this special section directly address these questions. Ronald, Happe, Dworzynski, Bolton, and Plomin (2010) provide evidence for the very weak effects of both prenatal and neonatal complications on the incidence of autistic like features in young children. Schmidt, Miskovic, Boyle, and Saigal (2010) report on the consequences, in adult life, of being born with very low birth weight in a nonmedically comprised sample. A third article by Grosse (2010) examines the consequences of late-treated phenylketonuria on cognitive development.

Much of the work on the effects of early experience and sensitive periods on later development has been with rodents or nonhuman primates. Data in human infants have been primarily in visual perceptual development (e.g., the work of Maurer, Lewis, Brent, & Levin, 1999) or auditory perceptual development and language (e.g., the work of Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Neville & Bavelier, 1998). Maurer's research examined perceptual development of infants born with bilateral cataracts, studying the effect of age at which the infant had surgery to remove the cataract on subsequent perceptual processes. Neville's work examined changes in brain organization and language abilities in infants born deaf. Kuhl showed how exposure to one's native language in the first months of life determines perceptual abilities to distinguish among different phonemic contrasts during later infancy and childhood. In the case of Maurer or Neville's work, use of subjects whose sensory faculties (either vision or hearing) were severely impaired allowed exploration of the effects of early experience and timing of those experiences on development. Examination of the effects of more general psychosocial deprivation and neglect in human infants has until recently been missing from the research field.

Studies of infants who have been abandoned at birth and placed into institutions where they received minimum basic care but have been deprived of typical psychosocial stimulation have filled the gap in our knowledge about the effects of these experiences on later cognitive and social development. These studies have been in the literature for some time (e.g., Tizard & Rees, 1974) but only recently has there been a systematic attempt to examine issues of timing and sensitive periods across a wide array of domains of cognitive and socioemotional behavior. A pioneering study in the United Kingdom (Rutter, O'Connor, and the ERA, 2004) tracked the cognitive and social development of a large group of children who were abandoned at birth into the institutions that existed in Romania. A series of articles from this study illustrated the effects of early severe psychosocial deprivation and neglect on children's subsequent development—as well as the importance of timing of intervention. Three articles in the current special section follow in this model: Pollak et al. (2010) present data from a large study of postinstitutionalized children adopted by families in the United States; Dobrova-Krol, van IJzendoorn, Bakermans-Kranenburg, and Juffer (2010) examine the consequences of institutionalization and HIV infection in infants in the Ukraine, and Smyke, Zeanah, Fox, Nelson, and Guthrie (2010) present evidence on the effects on social development of a foster care/family intervention within the context of the Bucharest Early Intervention Project. Of course, one does not have to study children in institutions to examine the effects of abuse and neglect on development. There is a long history of research into the consequences of physical and sexual abuse on children's development, exemplified in the current issue by a study by Cicchetti, Rogosch, Gunnar, and Toth (2010).

Two articles in this special section examine the effects of different early experiences on children's stress reactivity. Obradović, Bush, Stamperdahl, Adler, and Boyce (2010) use a model of biological sensitivity to context in which they examine whether certain contexts and early adversity are actually advantageous for some children while detrimental to others. In a similar vein, O'Neal et al. (2010) study the effects of a randomized clinical intervention trial aimed at reducing child aggression and stress reactivity. Their findings illustrate that often one particular intervention does not affect all children in the same way.

In the sixth part of this special section three articles illustrate both the mechanism and the importance of individual differences in temperament and genetics in predicting cognitive and socioemotional development. Bernier, Carlson, and Whipple (2010) nicely demonstrate how differences in maternal behavior can enhance the emergence of executive functioning in young children. This article serves as a reminder that brain areas continue to mature well into childhood and possibly adolescence, and that experiences can enhance the emergence of functions that are subsumed by these important late maturing brain regions. Leve et al. (2010) using a longitudinal design, including a prospective adoption cohort, examine both the genetic and caregiving effects predicting externalizing behaviors. Their findings illustrate both the unique and interactive effects of maternal caregiving and infant genetic makeup in predicting behavioral outcomes. The third article in this section (Duncan, Ziol-Guest, & Kalil, 2010) examines data from the Panel Study of Income Dynamics assessing the long-term effects of poverty on adult achievement, health, and behavioral outcomes. The study confirms the fact that early experience of poverty (in this case between children's prenatal to 5th year of life) has long-term consequences on a number of health and income outcomes.

The final article of this special section is a commentary by Shonkoff (2010) that lays out the relations between basic research into the effects of early experience and constructive social policy for the welfare and good of children. Shonkoff argues that the scientific evidence—as presented in the articles in this special issue—warrants innovative approaches to creating effective and novel prevention and intervention programs for children. He presents arguments for new strategies that can reduce the effects of disadvantage and early experience with "toxic" stress. His commentary demonstrates the importance of integrating established scientific evidence from brain and behavioral sciences, benefit–cost data, and the results of randomized controlled trials on intervention/prevention. It provides the link between the articles in this special section to important policy initiatives for the good of all children.

References

Bernier, A., Carlson, S. M., & Whipple, N. (2010). From external regulation to self-regulation: Early parenting precursors of young children's executive functioning. *Child Development*, *81*, 326–339.

Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., et al. (2003). Influence of lifestress on depression: Moderation by a polymorphism in the 5-HTT gene. *Science*, *301*, 386–389.

Cicchetti, D., Rogosch, F. A., Gunnar, M. R., & Toth, S. L. (2010). The differential impacts of early physical and sexual abuse and internalizing problems on daytime cortisol rhythm in school-aged children. *Child Development*, *81*, 252–269.

Davis, E. P., & Sandman, C. A. (2010). The timing of prenatal exposure to maternal cortisol and psychosocial stress is associated with human infant cognitive development. *Child Development*, *81*, 131–148.

Denenberg, V. (1964). Critical periods, stimulus input, and emotional reactivity: A theory of infantile stimulation. *Psychological Review*, *71*, 335–351.

DiPietro, J., Kivlighhan, K., Costigan, K., Rubin, S., Shiffler, D., Henderson, J., et al. (2010). Prenatal antecedents of newborn neurological maturation. *Child Development, 81*, 115–131.

Dobrova-Krol, N. A., van IJzendoorn, M. H., Bakermans-Kranenburg, M. J., & Juffer, F. (2010). Effects of perinatal HIV infection and early institutional rearing on physical and cognitive development of children in Ukraine. *Child Development, 81*, 237–251.

D'Onofrio, B., Singh, A. L., Iliadou, A., Lambe, M., Hultman, C., Neiderhiser, J. M., et al. (2010). A quasi-experimental study of maternal smoking during pregnancy and offspring academic achievement. *Child Development, 81*, 80–100.

Duncan, G. J., Ziol-Guest, K. M., & Kalil, A. (2010). Early childhood poverty and adult attainment, behavior, and health. *Child Development, 81*, 306–325.

Fox, S., Levitt, P., & Nelson, C. A. (2010). How the timing and quality of early experiences influence the development of brain architecture. *Child Development, 81*, 28–40.

Granger, B., Tekaia, F., Le Sourd, A. M., Rakic, P., & Bourgeois, J. P. (1995). Tempo of neurogenesis and synaptogenesis in the primate cingulate mesocortex: Comparison with the neocortex. *Journal of Comparative Neurology, 360*, 363–376.

Greenough, W. T., Black, J. E., & Wallace, C. S. (1987). Experience and brain development. *Child Development, 58*, 539–559.

Grosse, S. (2010). Late-treated phenylketonuria and partial reversibility of intellectual disability. *Child Development, 81*, 200–211.

Hay, D., Pawlby, S., Waters, C. S., Perra, O., & Sharp, D. (2010). Mothers' antenatal depression and their children's antisocial outcomes. *Child Development, 81*, 149–165.

Hess, E. H. (1964). Imprinting in birds: Research has borne out the concept of imprinting as a type of learning different from association learning. *Science, 146*, 1128–1146.

Hubel, D. H., & Wiesel, T. N. (1970). The period of susceptibility to the physiological effects of unilateral eye closure in kittens. *Journal of Physiology (London), 206*, 419–436.

Huttenlocher, P. R., & Dabholkar, A. S. (1997). Regional differences in synaptogenesis inhuman cerebral cortex. *Journal of Comparative Neurology, 387*, 167–178.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255*, 606–608.

Leve, L., Kerr, D., Shaw, D., Ge, X., Neiderhiser, J., Scaramella, L., et al. (2010). Infant pathways to externalizing behavior: Evidence of genotype x environment interaction. *Child Development, 81*, 340–356.

Levine, S. (1957). Infantile experience and resistance to physiological stress. *Science, 126*, 405.

Maurer, D., Lewis, T. L., Brent, H. P., & Levin, A. V. (1999). Rapid improvement in the acuity of infants after visual input. *Science, 286*, 108–110.

Meaney, M. J. (2001). Maternal care, gene expression, and the transmission of individual differences in stress reactivity across generations. *Annual Review of Neuroscience, 24*, 1161–1192.

Meaney, M. (2010). Epigenetics and the biological definition of gene × environment interactions. *Child Development, 81*, 41–79.

Neville, H., & Bavelier, D. (1998). Neural organization and plasticity for language. *Current Opinions in Neurobiology*, 254–258.

Obradović, J., Bush, N. R., Stamperdahl, J., Adler, N. E., & Boyce, W. T. (2010). Biological sensitivity to context: The interactive effects of stress reactivity and adversity on socioemotional behavior and school readiness. *Child Development, 81*, 270–289.

O'Neal, C. R., Brotman, L. M., Huang, K., Gouley, K. K., Kamboukos, D., Calzada, E. J., et al. (2010). Understanding relations among early family environment, cortisol response, and child aggression via a prevention experiment. *Child Development, 81*, 290–305.

Pollak, S. D., Nelson, C. A., Schlaak, M. F., Roeber, B. J., Wewerka, S. S., Wiik, K. L., et al. (2010). Neurodevelopmental effects of early deprivation in post-institutionalized children. *Child Development, 81*, 224–236.

Price, T. S., Grosser, T., Plomin, R., & Jaffee, S. R. (2010). Fetal genotype for the xenobiotic metabolizing enzyme *NQO1* influences intrauterine growth among infants whose mothers smoked during pregnancy. *Child Development, 81*, 104–114.

Ronald, A., Happe, F., Dworzynski, K., Bolton, P., & Plomin, R. (2010). Exploring the relationship between prenatal and neonatal complications and later autistic-like features in a representative community sample of twins. *Child Development, 81*, 166–182.

Rutter, M. (2007). *Genes and behavior: Nature–nurture interplay explained*. New York: Wiley.

Rutter, M., O'Connor, T. G., & the ERA. (2004). Are there biological programming effects for psychological development? Findings from a study of Romanian adoptees. *Developmental Psychology, 40*, 81–94.

Schmidt, L., Miskovic, V., Boyle, M., & Saigal, S. (2010). Frontal EEG asymmetry, salivary cortisol, and internalizing behavior problems in young adults who were born at extremely low birth weight. *Child Development, 81*, 183–199.

Shonkoff, J. (2010). Building a new biodevelopmental framework to guide the future of early childhood policy. *Child Development, 81*, 357–367.

Smyke, A., Zeanah, C. H., Fox, N. A., Nelson, C. A., & Guthrie, D. (2010). Placement in foster care enhances quality of attachment among young institutionalized children. *Child Development, 81*, 212–223.

Tizard, B., & Rees, J. (1974). A comparison of the effects of adoption, restoration to the natural mother, and continued institutionalization on the cognitive development of four-year-old children. *Child Development, 45*, 92–99.

Child Development, January/February 2010, Volume 81, Number 1, Pages 28–40

How the Timing and Quality of Early Experiences Influence the Development of Brain Architecture

Sharon E. Fox
Harvard–MIT Division of Health Sciences and Technology

Pat Levitt
Vanderbilt University

Charles A. Nelson III
Children's Hospital Boston, Harvard Medical School

Early life events can exert a powerful influence on both the pattern of brain architecture and behavioral development. In this study a conceptual framework is provided for considering how the structure of early experience gets "under the skin." The study begins with a description of the genetic framework that lays the foundation for brain development, and then proceeds to the ways experience interacts with and modifies the structures and functions of the developing brain. Much of the attention is focused on early experience and sensitive periods, although it is made clear that later experience also plays an important role in maintaining and elaborating this early wiring diagram, which is critical to establishing a solid footing for development beyond the early years.

Our current understanding of the role of experience in influencing the course of human development arises in large part from studies of atypical patterns and disorders of development, including disorders that arise due to adverse early experience. However, whereas studies of atypical development have done an excellent job of *describing* dimensional phenotypes that underlie disrupted development, it is imperative that we integrate these descriptive concepts with biological constructs of neurodevelopment in order to understand the relevant brain architecture and neurochemical constituents that determine responses to experience that ultimately lead to typical and atypical change (see Nelson & Jeste, 2008, for recent review). In this study we seek to elucidate the myriad ways the structure of experience weaves its way into the structure of the developing brain.

Background

The foundations of brain architecture are established early in life through a continuous series of dynamic interactions between genetic influences and environmental conditions and experiences (Friederici, 2006; Grossman, 2003; Hensch, 2005; Horn, 2004; Katz & Shatz, 1996; Majdan & Shatz, 2006; Singer, 1995). There is increasing evidence that environmental factors play a crucial role in coordinating the timing and pattern of gene expression, which in turn determines initial brain architecture. Because specific experiences potentiate or inhibit neural connectivity at key developmental stages, these time points are referred to as sensitive periods (Hess, 1973; Knudsen, 2004). Each one of our perceptual, cognitive, and emotional capabilities is built upon the scaffolding provided by early life experiences. Examples can be found in both the visual and auditory systems, where the foundation

---

The writing of this article was made possible by the Doris Duke Charitable Foundation Fellowship at Harvard Medical School (to Sharon Fox), the NIH (NS03445; MH078829), the John D. and Catherine T. Macarthur Foundation and the Richard David Scott Chair (to Charles A. Nelson), and the Annette Schaffer Eskind Chair, NIMH Grant MH080759 and NICHD P30 core Grant HD15052 (to Pat Levitt).

Correspondence concerning this article should be sent to Charles A. Nelson, Children's Hospital Boston, Harvard Medical School, 1 Autumn Street, 6th Floor, Mailbox #713, Au621, Boston, MA 02215-5365. Electronic mail may be sent to charles.nelson@childrens.harvard.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8101-0004

for later cognitive architecture is laid down during sensitive periods for basic neural circuitry. The capacity to perceive stereoscopic depth requires early experience with binocular vision (Crawford, Pesch, & von Noorden, 1996), which at a later point in development may have implications for perceptual and cognitive development. Likewise, the capacity to perceive a range of tones requires variation in the tonal environment, and exposure to such variation later leads to language processing and proficiency (Kuhl, 2004; Newport, Bavelier, & Neville, 2001; Oyama, 1990; Weber-Fox & Neville, 2001). Just as a square house cannot be built on a round foundation, so must later refinements in adult perceptions reflect the architectural foundation developed through early experiences. Although early experiences are reflected in behavior, behavioral measures tend to underestimate (in part because of a lack of sensitivity and specificity) the magnitude and persistence of the effects of early neuronal development (Knudsen, 2004). In order to explore the role of timing and quality of early experiences on later cognitive function, we must therefore return to the genetic framework of the developing brain.

Genetics and Environment—The "Blueprint" for Neural Development

Refinements in the neural circuits that mediate sensory, emotional, and social behaviors are driven by experience (Feldman & Knudsen, 1998). Specifically, postnatal experiences drive a protracted process of maturation at the structural and functional level, but the very ability of such developmental processes to occur successfully is dependent in large part on the prenatal establishment of the fundamental brain architecture that provides the basis for receiving, interpreting, and acting on information from the world around us (Hammock & Levitt, 2006). While the term *blueprint* has been utilized in the past to describe a fixed set of genes with inflexible interactions, the term is used here as an analogy to a rough draft, or design—the framework from which a more defined structure will evolve. The emergence of the architecture in all vertebrate species begins early, when the fundamental cardinal axes are established in the neural plate by sets of different transcription factors that specify molecular differences, particularly growth factor signaling molecules (e.g., fibroblast growth factors, wnts) antero-posteriorly, medio-laterally, and dorso-ventrally; Puelles, 2001). In humans, this occurs within the first 2 months postfertilization (Levitt, 2000, 2003). The initial focus of brain patterning was on the nature of specification of segmented regions of the nervous system, such as the brainstem and spinal cord, in which highly conserved homeobox-containing transcription factors control "downstream" gene networks that specify motor and sensory neuron types in each segment. It has become clear, however, that developmental gene networks are a highly conserved strategy for specifying the initial plan of the central nervous system. Thus, they are utilized even in the more complex, nonsegmented anterior regions of the brain that include the thalamus, hypothalamus, basal forebrain, and cerebral cortices (Sur & Rubenstein, 2005). The gene networks in large part encode transcription factors, proteins that control (through positive [enhancing] and negative [repressing] activity) the expression of genes downstream that mediate the major histogenic events that include progenitor cell proliferation, cell fate choices, cell migration, axon and dendritic growth, and synapse formation. Collectively, these highly structured events build the early plan for brain architecture. The genetic mechanisms are thought to be highly conserved because when examined, the networks are expressed in similar patterns across rodents and humans. Experimentally, however, the genetic mechanisms through which regional patterns and connections are established prenatally have been defined in animal models, primarily rodents and nonhuman primates (O'Leary, Chou, & Sahara, 2007; Rakic, 2006). In the context of understanding clinical disorders that have a prenatal origin, the challenge has been to extrapolate that information from different species in the context of the timing of human brain development, an exercise that has received considerable attention with new database tools (e.g., http://www.translatingtime.net/). The onset of histogenesis during prenatal development provides the cellular framework necessary for establishing and modifying later developing circuitry.

The cerebral cortex has garnered substantial attention in defining key developmental features across species. This is due in part to the technical advantages of studying a well-organized, layered structure, and the functional relevance of linking typical and atypical maturation of complex behaviors and neurodevelopment. The neocortex in all mammalian species comprises six layers of neurons, the architecture, connectivity, and chemistry of which are distinct depending upon their location. The neocortex is organized to receive information from the organism's surrounding environment, typically through connections with the thalamus. It does so by integrating information within and across architecturally distinct functional domains, and then relays this information to other brain centers that generate an appropriate functional response. There are two major organizing principles of the neocortex that are guided by gradients of expressed gene networks that establish an evolutionarily conserved design. First, the precursors of different functional areas emerge during the time period when neurons are produced and are displayed in the tangential domain (roughly the first and second trimesters of pregnancy in the human; see O'Leary et al., 2007; Sur & Rubenstein, 2005). Regional specification is not absolute but does involve transcriptional networks controlling the expression of axon guidance molecules that control the initial input and output wiring plan (see below). Expansion of the size of the neocortex during evolution (e.g., 1,000-fold between mouse and human) occurs mostly in this tangential domain (Rakic, 2005). Extending through the six-layered neocortex is the most fundamental of functional units, the radial column (Rakic, 2006). Each column of cells contains well-organized networks of local circuit and projection neurons, and are built in an orderly fashion in which the neurons destined to occupy deep layers are produced, or ''born'' first, followed by more superficially displaced neurons. Though there are some differences between rodents and primates, the projection neurons, which utilize the excitatory neurotransmitter glutamate, and the local circuit neurons, many of which use the inhibitory neurotransmitter gamma-amino butyric acid (GABA), arise in part from different regions of the forebrain (Wonders & Anderson, 2007). The regions of the forebrain proliferative zones from which glutamatergic and GABAergic neurons arise are specified genetically to give rise to the unique neuron types by different sets of transcription factors. For example, glutamatergic neurons arise from the dorsal pallium, the roof of the forebrain in which cells uniquely express transcription factors like Emx1 and Pax6. GABAergic neurons destined for the cerebral cortex, joining their glutamateric neuron partners, arise from the subpallium which also gives rise to the basal ganglia. This region expresses transcription factors in the Dlx and Nkx family. The genetic influence of these initial neurodevelopmental events is illustrated by experiments in which these different transcription factors are misexpressed in different regions will result in the anomalous production of projection or interneurons in the wrong location (Hebért & Fishell, 2008).

The ''inside-out'' pattern of neuron production and migration provides the basis for building essentially similar ontogenetic radial units across functional areas, with minor variations in the ratio of excitatory to inhibitory neurons in different regions. In fact, this conserved radial organization provides the framework for later developing refinement of circuits that are influenced extensively by patterns of physiological activity through experience-dependent mechanisms (see below). The tangential patterning of the neocortex into different functional domains that are prepared to receive specific thalamic input begins as soon as the first neurons are produced and is due largely to distinctive gradients of expression of transcription factors such as emx1 and 2, pax6 and lhx2, which regulate the production of signaling molecules that include the fibroblast growth factors (fgf), bone morphogenetic proteins (bmp), and wnt proteins (O'Leary et al., 2007). Experiments in genetically manipulated mice demonstrate that by eliminating or altering the expression of just one transcription factor, the functional fate of cortical regions can be changed (Cholfin & Rubenstein, 2007). For example, *emx2* controls the expression of fgf8 near the anterior end of the cerebrum. Fgf8 alone is sufficient to specify the cortical regions that will eventually receive connections that are typical of frontal and somatosensory cortices (Fukuchi-Shimgori & Grove, 2001, 2003). This type of early genetic respecification is functionally relevant. For example, Fgf17 is responsible for the initial patterning of different frontal cortex areas (Cholfin & Rubenstein, 2007). The early specification of the neocortex by genetic mechanisms is powerful because downstream from these signaling factors are axon guidance molecules that serve as the important chemical cues for getting axons to grow into their correct target region prior to beginning the extended process of synapse formation (e.g., see Alcamo et al., 2008). Gene regulatory networks also can influence the initial size of cortical areas by modulating the number of neurons that are produced. As noted above, key to the flexibility of this early architecture is the conserved nature of the radial organization. Thus, the long-distance circuit projections that help to define functional cortical areas, and even functional differences in superficial and deep projecting neurons, are altered when the disruption of early gene networks modifies guidance cues so that atypical connections are made. Though we tend to think that genetic mechanisms are immutable, it is important to stress that expression of early gene networks can be perturbed not only by catastrophic genetic mutations that disrupt important regulatory genes but also by prenatal environmental influences, such as drugs, alcohol, toxins, and inflammatory responses. These may have less profound impacts on brain patterning but nonetheless can result in long-term disruption of cellular differentiation and behavioral development (Stanwood & Levitt, 2008).

In all mammalian species, this early period of specified patterning to generate a unique architecture is followed by an extended period of synapse formation, adjustment and pruning that typically extends from the last quarter of gestation through puberty (Bourgeois, Goldman-Rakic, & Rakic, 1999; Huttenlocher & Dabholkar, 1997). There is a unique relation that emerges between pre- and postsynaptic partners as physiological activity takes over from the activity-independent histogenic events to sculpt presynaptic terminals and postsynaptic dendrites and spines. Over the past decade, investigators have identified literally thousands of genes that encode proteins related to synapse formation, plasticity, and stabilization (Akins & Biederer, 2006; Sheng & Hoogenraad, 2007; Sudhof, 2008). While synapses can form under experimental conditions in the absence of physiological activity, experience is essential for the normally occurring regulation of the molecular basis for synapse formation. In this complex process, activity controls the expression of transcription factors that direct the downstream expression of structural proteins, receptors, and signaling molecules that are needed for synapses to function properly (Flavell et al., 2006; Majdan & Shatz, 2006; Paradis et al., 2007; Sugiyama et al., 2008). Moreover, activity regulates the distribution of key proteins within the synapse, making them available for the important task of information processing during sensitive and critical periods of development (Shepherd & Huganir, 2007).

Because of dramatic differences in brain size, the developmental timing of the histogenic events across mammalian species is unique. The process begins at about 56 days postconception and extends through the first years of life in humans. The events (e.g., cell type specification, axon guidance) that provide the initial brain blueprint, influenced by early genetic patterning, are completed early and relatively rapidly, by midgestation in the primates and the end of gestation in rodents (Levitt, 2003). The formation of neural connections, the development of unique cytoarchitecture (the structural correlate of functional areas), the growth of dendritic arbors, and the peak formation of synapses is a far more time-extensive process, extending through the second and third postnatal years in humans, 180 days in the macaque monkey, and at around weaning (21 days) in rodents. Thus, through a well-defined developmental process, the basic connectivity of local and long-distance neurons is set up to take advantage of a highly flexible organization that is both activity dependent and expectant postnatally. The continued growth of the neocortex occurs up to puberty by the addition of myelin, dendritic growth, non-neuronal cells, and a complex process of resculpting synapses whose numerical density across the entire neocortex is stable but may change in their laminar distribution. The mechanisms that control this latter event are not clear but are likely to be activity dependent. Recent studies of synaptic resculpting demonstrate that highly novel gene–environment regulatory mechanisms are at play, such as the activity-dependent movement of transcription factors from pre- to postsynaptic neurons to control maturation of neurons in a noncell autonomous fashion (Sugiyama et al., 2008).

Developmental Plasticity—Adjustments to the Architectural Plan

Although our genetic code provides an important foundation for early development, it must be understood as a framework upon which many environmental factors influence future structure and function. This can best be illustrated through studies of the sensory systems, which demonstrate the crucial role of environment in the early development and maintenance of the nervous system. Such work also demonstrates the need for patterned physiologic activity during development, as well as refinement and maintenance of detailed sensory maps. For example, supporting cells in the developing rat cochlea spontaneously release adenosine-5′-triphosphate (ATP), synchronizing the output of neighboring inner hair cells, which aids in the refinement of tonotopic maps. Spontaneous ATP-dependent signaling rapidly subsides after the onset of hearing, thereby preventing this experience-independent activity from interfering with the accurate encoding of sound. The initiation of electrical activity in auditory nerves before hearing suggests that peripheral, nonsensory cells may provide the necessary early environment for the development of central auditory pathways (Tritsch, Yi, Gale, Glowatski, & Bergles, 2007). In addition, auditory perception evolves according to the tonal rearing environment, with increased sensitivity to tones outside of those continuously experienced in the environment (Hanm, Köver, Insanally, Semerdjian, & Bao, 2007).

Similarly, in the landmark studies of vision by Wiesel and Hubel (1963, 1965), it was demonstrated that kittens reared with normal visual experience resulted in each eye having sole access to alternating columns of neurons in layer IV of the striate cortex. At birth, however, both eyes synapse on *all* neurons in layer IV. In order to assure that a neuron is stimulated by experience coming from only one eye, a competitive process occurs in which activation and neighboring inhibition result in an alternating pattern of connectivity between columns of neurons in layer IV and each eye (Wiesel & Hubel, 1965). When kittens were reared with one eye closed for a period of time after birth, the occluded eye became essentially functionally blind. This blindness is due to the elimination of connections of the closed eye to layer IV and the lack of exposure to patterned activity. If occlusion extends beyond a certain time period, the typical pattern of ocular representation cannot be recovered despite the restoration of visual input to both eyes (Wiesel & Hubel, 1965). It has been hypothesized that the initial ingrowth of axons from the thalamus to ocular dominance columns in the visual cortex is governed by molecular cues (Crair, Horton, Antonini, & Stryker, 2001; Crowley & Katz, 2000). That is not to say, however, that an *atypical* or *plastic* change in visual representation is not achievable (Amedi et al., 2007; Antonini, Fagiolini, & Stryker, 1999; Flege, Munro, & MacKay, 1995; He, Ray, Dennis, & Quinlan, 2007; Miller, Keller, & Stryker, 1989). It has recently been shown, for example, that the decreased visual acuity seen in the adult rat suffering from chronic monocular deprivation is reversed if the adult rat is treated with dark exposure prior to removal of the occlusion (He et al., 2007). The increased plasticity induced by the dark environment may be due to a lack of input to the visual cortex through the functioning eye, and therefore a reduction in the strength of previously established connections. A similar restoration of visual acuity can also be induced with chronic administration of fluoxetine (Maya Vetencourt et al., 2008). Such dramatic changes in sensory system connectivity suggest that activity-dependent potentiation of these initial axons is required to maintain connections among cortical regions. In the case of the primary visual cortex, local circuit neurons have been implicated in activity-dependent plasticity through GABAergic inhibition over a wide range of neighboring axonal paths. (Fagiolini et al., 2004; Hensch & Stryker, 2004). An altered pattern of activity through one circuit can thus radically change neighboring circuits through an increase or decrease in inhibition of mediating cells. Furthermore, visual deprivation has been shown to alter the formation of dendritic spines in the visual cortex once the pathways for normal vision have begun to potentiate (Oray, Majewska, & Sur, 2004; Trachtenberg & Stryker, 2001). Following deprivation, spines show increased structural motility—possibly allowing for plasticity of local connections—and later deterioration (Kim & Bonhoeffer, 1994; Mataga, Mizuguchi, & Hensch, 2004).

The early development of visual pathways may be likened to the laying of a foundation and scaffolding for a building. If the scaffolding pattern is changed, the building may not be constructed in its original form, though a functional alternative may be reached. Thus, irreversible changes at the synaptic level do not necessarily translate into irreversible changes in a complex behavior (Feldman & Knudsen, 1998). For example, we now understand that the sensitive period for visual representation reflects, predominantly, the critical period for thalamic input to layer IV (Antonini & Stryker, 1993; Miller et al., 1989; Pascual-Leone, Amedi, Fregni, & Merabet, 2005), but that plasticity of other sensory systems may allow a blind person to demonstrate normal—and possibly enhanced—spatial awareness (Amedi et al., 2007). Plasticity in higher regions involved in spatial awareness feeds back upon lower pathways, thus compensating for an abnormal visual representation.

Advanced perceptual processes are also dependent upon the normal development of basic visual systems. For example, early visual deprivation due to congenital cataracts can lead to subtle but persistent deficits in face processing, even when the cataracts are removed in the first months of life (Le Grand, Mondloch, Maurer, & Brent, 2001). Similarly, experience with specific faces, such as same versus different species, powerfully shapes subsequent face specialization. For example, monkeys deprived of viewing faces since birth are capable of discriminating both monkey and human faces following the selective restoration of faces in the visual environment, but *what kind* of faces determines whether that same monkey will be able to subsequently discriminate human or monkey faces—thus, monkeys selectively exposed to human faces can only discriminate human faces, not monkey faces, and monkeys selectively exposed to monkey faces can only discriminate monkey faces, not human faces (Sugita, 2008).

Sensitive Periods in the Early Development of Higher Cognitive Processes

Wiesel and Hubel's (1963, 1965) experiments involving visual deprivation brought about the concept of "sensitive" and "critical" periods in early cognitive development. "Sensitive" periods are defined as a time in development during which the brain is particularly responsive to experiences in the form of patterns of activity (Daw, 1997). Further, this time point may be termed a "critical" period if the presence or absence of an experience results in irreversible change (Newport et al., 2001; Trachtenberg & Stryker, 2001). Those factors that allow a circuit underlying cognition to be plastic—or render it unchangeable—are not yet well understood. In the area of speech and language, the maturational hypothesis predicts that native language proficiency cannot be obtained when learning begins after puberty (Bruer, 2001; Werker & Tees, 2005). Studies supporting this theory have correlated the degree of accent in a second language with age at the time of acquisition of that language (Birdsong & Molis, 2001; Johnson & Newport, 1989). Adults exposed to a second language in early childhood were found to have native-like accents and pattern of tone (Gordon, 2000; Long, 1990; Oyama, 1990; Stein et al., 2006). Other researchers have also found a negative correlation between age at acquisition and grammaticality judgments (Flege, MacKay, & Meador, 1999; Komarova & Nowak, 2001). These same studies, however, have failed to show any clear discontinuity in the relation between accent and age over the period from childhood to adulthood, thus suggesting that although a sensitive period may exist, a critical period for language acquisition is unlikely (Flege et al., 1999; Komarova & Nowak, 2001; Wartenburger et al., 2003; Weber-Fox & Neville, 2001). Many of these same investigators, however, argue that the decline in language-learning facility may be confounded by educational level and that involvement in a larger bilingual community may diminish the drive to find means for communication through a learned language (Flege et al., 1999; Hoff & Naigles, 2002; Komarova & Nowak, 2001). It has further been hypothesized that enhancements in nonlanguage cognitive systems, such as memory, may actually have deleterious effects on language learning (Newport, 1990; Newport et al., 2001). If this is the case, then the neural system responsible for language may in fact remain open, though difficult to access through adult cognitive processing or with only typical effort (the extent to which considerable effort might force open the sensitive period can be found in training studies such as those focused on teaching native Japanese speakers to perceptually discriminate and correctly produce the R/L distinction (see Akahane-Yamada, Strange, Downs-Pruitt, & Masuda, 1998; Guion, Flege, Akahane-Yamada, & Downs-Pruitt, 1998; Guion, Flege, Akahane-Yamada, & Pruitt, 2000).

Several investigators have used the theory of neural networks, originally developed for vision research, to model the activity of individual neurons and/or groups of neurons in the brain during learning (Christiansen & Chater, 2001; Morton & Munakata, 2005). These neural network models are particularly useful for comparing the experience-independent and experience-based accounts of sensitive periods because the network can be kept constant with regard to features affected by maturation, motivation, and amount of exposure. Returning to the work of Wiesel and Hubel (1963, 1965), it is important to note that the loss of binocular function in the kitten did not arise simply because of the absence of input to the occluded eye. Occluding both eyes during the same time period of development was proved not to result in loss of binocular vision (Cynader & Mitchell, 1980). It is necessary for one eye to have access to layer IV of the visual cortex whereas the other eye is denied access, allowing exclusive connectivity of the unoccluded eye to striate cortex. The irreversible loss of binocular vision during development must therefore be due to a combination of environmental experience and cortical learning processes (Knudsen, 2004; Mataga et al., 2004). The fact that the existence of a sensitive period can depend upon occurrence of a particular environment suggests that in early development, portions of networks become perceptually biased, making future modifications more difficult. For example, in the literature on both speech and face perception, the perceptual window through which faces and speech is initially processed is broadly tuned, then narrows with experience. For example, Pascalis, de Haan, and Nelson (2002) demonstrated that 6- and 9-month-old infants and adults can readily discriminate two human faces, but only 6-month-old infants can discriminate two monkey faces. Similarly, 6-month-olds given 3 months of experience viewing monkey faces can readily discriminate monkey faces at 9 months of age, whereas 9-month-old infants not afforded such experience cannot (Pascalis et al., 2005).

As a rule, circuits that process lower level information mature earlier than those that process higher level information (Burkhalter, Bernardo, & Charles, 1993; Scherf, Behrmann, Humphreys, & Luna, 2007). For example, in the neural hierarchy that analyzes visual information, low-level circuits that analyze the color, shape, or motion of stimuli are fully mature long before the high-level circuits that analyze or identify biologically important stimuli, such as faces, food, or frequently used objects (Burkhalter et al., 1993; Knudsen, 2004; Scherf et al., 2007). The process by which initial learning leads to a constraint on later learning is termed *entrenchment* and is equally apparent in the development of speech (Munakata & Pfaffly, 2004; Seidenberg & Zevin, 2006). Several studies have shown, for example, that adults are often better at discriminating nonnative phonetic contrasts when they differ substantially from phonemes of their native language (Best, McRoberts, & Sithole, 1988; Frieda, Walley, Flege, & Sloane, 1999; Guion et al., 2000; Kuhl, 2004). Adults are poorer at discriminating when the phonetic contrasts are similar to phonetic contrasts of their native language. This is akin to the nature of the developing auditory system, which as previously noted is more capable of discriminating tones outside of the tonal environment of rearing. At both the level of tone and of speech phonetic discrimination, there is evidence for a fixed bias of the neural network. As previously discussed in the case of visual networks, however, neurons may be constantly modifying connectivity, allowing learning from new environments to compete against already existing tendencies. This is well demonstrated in animals altered neonatally to receive retinal projections to the auditory portions of the thalamus (von Melchner, Pallas, & Sur, 2000). Such animals reveal that auditory cortex may be modified by extrinsic activity to develop retinotopic maps similar to those seen in the visual cortex. The role of environment and inputs to the brain may therefore be seen as critical in the bias of network formation during early life.

Altered patterns of enhancement and inactivity are thought to be the basis for neural plasticity and have been suggested in humans by studies of tactile and auditory perception in the blind, where such systems may even activate the "visual" cortex (Antonini & Stryker, 1993; Ellis & Lambon, 2000; Merabet, Rizzo, Amedi, Somers, & Pascual-Leone, 2005). It is likely that changes in experience have a greater impact on an untrained "young" network compared to the same experience on an "older" trained network. This biasing feature is suggested by studies on aphasia that show that words learned earlier in life are more resistant to loss and are more easily accessed in naming tasks compared to words learned later (Greenough, Black, & Wallace, 1987).

It has been suggested that learning through experience leads to the capacity to understand specific environments and the responses needed for these environments (Anisman, Zaharia, Meaney, & Merali, 1998; Scarr & McCartney, 1983). Similarly, changes in the environment—particularly when they are dramatic and pervasive—may have the power to alter neural connectivity and cognitive processing between systems. Examples can be found in studies of sensory deprivation, such as blindfolding, as well as sensory enhancement through technology. In studies of deaf children receiving cochlear implants, it is clear that language learning improves with earlier correction (Tomblin, Barker, Spencer, Zhang, & Gantz, 2005). It remains to be determined, however, whether this effect upon learning is due to actual changes in cognitive capacity or changes in the learning environment brought about by the ability to interact with others through spoken language.

Global Development—Higher Level Functions Build on Lower Level Functions

The nature of our experiences, particularly during a time-limited period in early development, can profoundly affect the mental framework we use to understand the world around us. Sensitive periods in child development are of interest because they represent a time frame in which our capabilities can be modified and perhaps enhanced. The quality of experiences during such periods—be they adverse or enhancing—is also of importance in understanding why it may be difficult to restore normal function once development has been altered. While explanatory models for the timing of early experiences have generally been based at the genetic or neural circuit level, our direct observations of the effects of early environments are often made at the behavioral level. Through the study of sensitive periods, we are better able to understand the impact that early experience may have upon development. To cite but one example, it has recently been demonstrated that otherwise typically developing young children institutionalized at birth have intelligence quotients (IQs) in the low 70s. However, placing such children in high-quality foster care *before the age of 2 years* leads to a dramatic increase in IQ (Nelson et al., 2007). A similar trend also occurs for language (Windsor, Glaze, Koga, &

the BEIP Core Group, 2007) and the development of the EEG (Marshall, Reeb, Fox, & the BEIP Core Group, 2008), although in the case of the former, the sensitive period occurs around 16–18 months.

It is important to note recent work suggesting that the brain retains the capacity to adapt and change throughout the lifespan (Crawford et al., 1996; Jones, 2000; Keuroghlian & Knudsen, 2007). However, the foundation of brain architecture must lie in the early developmental years, and the influence of childhood environment is much more salient in such basic cognitive processes as sensory perception (Amedi et al., 2007; Antonini & Stryker, 1993; Grossman et al., 2003; Hensch, 2005; Hess, 1973; Karmarkar & Dan, 2006; Knudsen, 2004; Pascual-Leone et al., 2005). Each sensory and cognitive system reaches a unique sensitive period (Daw, 1997), and thus identical environmental conditions will result in very different cognitive and emotional experiences for a child, depending upon his or her age (Amedi et al., 2007; Jones, 2000; Trachtenberg & Stryker, 2001; Tritsch et al., 2007; for reference, see Bailey, Bruer, Symons, & Lichtman, 2001).

Behavioral analysis can demonstrate the value of early experiences in the development of the brain. It must be remembered, however, that information is processed in a series of networks, each reflecting the effects of environment at varying time points. Higher level processing may mask modifications in lower level networks (Daw, 1997; Feldman & Knudsen, 1998; Trachtenberg & Stryker, 2001). Thus, behavioral outcomes may be shaped by later experience, even though circuits at the lowest levels in a pathway remain irreversibly altered. In addition, studies of the plasticity of sensory processing reveal that similar information can be derived from alternative pathways (Akins & Biederer, 2006; Antonini & Stryker, 1993; Ellis & Lambon, 2000; von Melchner et al., 2000; Pascual-Leone et al., 2005). For example, when using sound devices to asses space, blind individuals have been shown to activate the lateral occipital cortex in the same manner as sighted individuals do through vision (Amedi et al., 2007). It has been suggested that loss sensory input—such as occurs in late blindness—may in fact lead to the unmasking and strengthening of alternative pathways stemming from multisensory integration regions of the brain (Pascual-Leone et al., 2005). These pathways may not only substitute for the original sensory inputs but may also enhance previously existing capabilities. This form of sensory enhancement can often be seen in the highly tuned auditory and tactile perception of blind.

High-level neural circuits that carry out sophisticated mental functions depend on the quality of the information that is provided to them by lower level circuits. Low-level circuits whose architecture was shaped by healthy experiences early in life provide high-level circuits with precise, high-quality information. High-quality information, combined with sophisticated experience later in life, allows the architecture of circuits involved in higher functions to take full advantage of their genetic potential. Thus, early learning lays the foundation for later learning and is essential (though not sufficient) for the development of optimized brain architecture. Stated simply, rich early experience must be followed by rich and more sophisticated experience later in life, when high-level circuits are maturing, in order for full potential to be achieved (DeBello & Knudsen, 2004; Karmarkar & Dan, 2006; Nelson, de Haan, & Thomas, 2006; Sabatini et al., 2007).

Concluding Comments

Although studies of the adverse effects of deprivation on brain development are powerful and compelling, they tell us little about the benefits of enrichment. Much of what we know about the impact of early experience on brain architecture comes from animal or human studies of deprivation. As we work to clarify further the patterns of genetic expression required for normal neural structure, we have also recognized that an optimal level of environmental input or "expectable" environment must exist in parallel. Increasing evidence suggests that this "expectable environment" of early development requires not only the variation in light necessary for vision, or the tones heard in a spoken language, but also the emotional support and familiarity of a caregiver (Nelson et al., 2007; Sánchez, Ladd, & Plotsky, 2001). It is important to emphasize that the well-documented negative impacts of deprivation on brain circuitry do not mean that *excessive* enrichment produces measurable enhancements in brain function. A small number of case reports exist in which neglected children with very little language experience in early childhood were given enriched language exposure in a protective environment (Curtiss, 1977; Itard, 1801/1932; Zingg, 1940). Longitudinal follow-up studies of these children demonstrated that after several years of language exposure, they were unable to achieve adult-level native language abilities. In animal models, rat pups deprived of maternal care were shown to have reduced hippocampal volume compared to pups with "enriched" maternal care (Bredy, Humpartzoomian, Cain, & Meaney, 2003). Placement of deprived pups into an "enriched care" environment resulted in learning and memory aptitude similar to high-care pups; however, hippocampal volume did not change, suggesting that plastic mechanisms early in life allow for alternative pathways to form typical behavior despite lasting structural deficits. Most recently, early intervention to correct a deeply impoverished early environment has been shown to greatly improve cognitive, linguistic, and emotional capabilities in humans (Ghera et al., 2009; Nelson et al., 2007; Windsor et al., 2007). Activity-dependent mechanisms of network formation, as described earlier, may be responsible for such changes when children are placed into a stimulating environment for learning and exploration. With continued research into the modification of sensitive periods, as well as the factors influencing cortical plasticity throughout life, we may remain optimistic about the possibility of recovery from early deprivation. This in turn may provide hope for children who lack the biological framework or necessary environment required for optimal neural and cognitive growth.

Finally, the possibility of cognitive and neural rehabilitation leads to theories of enrichment beyond the norm to a level of *enhanced* development. Educational and environmental enrichment of preschool children from impoverished economic settings has been shown to improve cognitive measures through early adulthood (Campbell, Pungello, Miller-Johnson, Burchinal, & Ramey, 2001; Martin, Ramey, & Ramey, 1990; see Farran, 2000, for review). Cognitive capabilities, however, may follow a pattern similar to the growth curve of the human body; that is, although it is possible to enhance the environment of a child to assume the pattern of the normal curve, it is not possible to exceed the predicted trajectory to a significant extent without causing some potential harm. This is suggested by large studies of children from varying socioeconomic status, which demonstrated an improvement in cognitive performance only in those born to a low socioeconomic class, with no significant difference between those of middle-class and high-income families (Jeffaris, Power, & Hertzman, 2002). If the possibility for enhancement exists, it is perhaps related to forms of enrichment that lie outside access to unlimited resources—that is, beyond the expectable environment. Creativity, for example, is a key component to enhanced cognitive functioning, yet we have not been able to define the neural processes or environmental attributes that can enrich this aspect of cognition, nor are there sure-fire ways of boosting creativity among the population at large. Similarly, exposure to art or music or great literature or horseback riding may not confer any evolutionary advantage (i.e., reproductive success), yet these activities may confer some advantage among certain strata of society. Thus, perhaps it would be useful to draw a distinction between *enrichment* as applied to those experiencing downward deviations from the expectable environment (such as those reared in situations of neglect or deprivation) and *enhanced enrichment* applied to those reared in typical (expectable) environments. Enrichment may lead to a restoration of typical development whereas enhanced enrichment may lead to exceeding the typical environment. Of course, a challenge here lies in accounting for individual differences, as some individuals have greater potential to benefit from art or music lessons than others. Individual heterogeneity may be under control of Gene × Gene, Gene × Environment, and Environment × Environment factors. For example, animal studies show that epigenetic mechanisms—whereby environmental factors and experiences early in life can permanently alter the genome of an individual through chemical modification—will impact long-term cognitive and social-emotional functioning (Szyf, McGowan, & Meany, 2008). The field awaits translation of this type of mechanism into human experiences. Finally, how might the field of developmental psychology benefit from advances being made in developmental neuroscience? First, given that our genome contains many fewer genes than we surmised even a decade ago (approximately 20,000), and given advances being made in the field of epigenetics, renewed attention should be paid to the origins and elaboration of complex human behaviors. Second, those working in the field of intervention should take stock in what is now known about neural plasticity; for example, it is quite possible that we could witness a revolution in new treatment approaches based on what we know about the malleability of the human brain. Finally, for the millions of children around the world who begin their lives in adverse circumstances, we should be mindful of what is known about sensitive periods and act with alacrity to improve the lives of these children before neural circuits become well established and, thus, difficult to modify. To borrow an analogy from economics, by investing early and well in our children's development, we increase the rate of return later in life and in so doing improve not only the lives of individuals but of societies as well.

References

Akahane-Yamada, R., Strange, W., Downs-Pruitt, J. C., & Masuda, Y. (1998). Modification of L2 vowel production by intensive perception training as evaluated by acoustic analysis and native speakers. *Journal of the Acoustical Society of America, 103*, 3089–3096.

Akins, M. R., & Biederer, T. (2006). Cell–cell interactions in synaptogenesis. *Current Opinion in Neurobiology, 16*, 83–89.

Alcamo, E. A., Chiriella, L., Dautzenberg, M., Dobreva, G., Fariñas, I., Grosschedl, R., et al. (2008). Satb2 regulates callosal projection neuron identity in the developing cerebral cortex. *Neuron, 57*, 364–377.

Amedi, A., Stern, W. M., Camprodon, J. A., Bermpohl, F., Merabet, L., Rotman, S., et al. (2007). Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nature Neuroscience, 10*, 687–689.

Anisman, H., Zaharia, M. D., Meaney, M. J., & Merali, Z. (1998). Do early-life events permanently alter behavioral and hormonal responses to stressors? *International Journal of Developmental Neuroscience, 16*, 149–164.

Antonini, A., Fagiolini, M., & Stryker, M. P. (1999). Anatomical correlates of functional plasticity in mouse visual cortex. *Journal of Neuroscience, 19*, 4388–4406.

Antonini, A., & Stryker, M. P. (1993). Rapid remodeling of axonal arbors in the visual cortex. *Science, 260*, 1819–1821.

Bailey, D. B., Bruer, J. T., Symons, F. J., & Lichtman, J. W. (Eds.). (2001). *Critical thinking about critical periods*. Baltimore, MD: Paul H. Brookes.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 345–360.

Birdsong, D., & Molis, M. (2001). On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language, 44*, 235–249.

Bourgeois, J.-P., Goldman-Rakic, P. S., & Rakic, P. (1999). Formation, elimination, and stabilization of synapses in the primate cerebral cortex. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 45–53). Cambridge, MA: MIT Press.

Bredy, T. W., Humpartzoomian, R. A., Cain, D. P., & Meaney, M. J. (2003). Partial reversal of the effect of maternal care on cognitive function through environmental enrichment. *Neuroscience, 118*, 571–576.

Bruer, J. T. (2001). A critical and sensitive period primer. In D. Ailer Jr., J. T. Bruer, F. J. Symons, & J. W. Lichtman (Eds.), *Critical thinking about critical periods* (pp. 3–26). Baltimore, MD: Paul H. Brookes.

Burkhalter, A., Bernardo, K. L., & Charles, V. (1993). Development of local circuits in human visual cortex. *Journal of Neuroscience*, 13(5), 1916–1931.

Campbell, F. A., Pungello, E. P., Miller-Johnson, S., Burchinal, M., & Ramey, C. T. (2001). The development of cognitive and academic abilities: Growth curves from and early childhood educational experiment. *Developmental Psychology, 37*, 231–242.

Cholfin, J. A., & Rubenstein, J. L. R. (2007). Patterning of frontal cortex subdivisions by fgf17. *Proceedings of the National Academy of Sciences of the United States of America, 104*, 7652–7657.

Christiansen, M. H., & Chater, N. (2001). Connectionist psycholinguistics: Capturing the empirical data. *Trends in Cognitive Sciences, 1*, 82–88.

Crair, M. C., Horton, J. C., Antonini, A., & Stryker, M. P. (2001). Emergence of occular dominance columns in cat visual cortex by 2 weeks of age. *Journal of Comparative Neurology, 430*, 235–249.

Crawford, M. L., Pesch, T. W., & von Noorden, G. K. (1996). Excitatory binocular neurons are lost following prismatic binocular dissociation in infant monkeys. *Behavioural Brain Research, 79*, 227–232.

Crowley, J. C., & Katz, L. C. (2000). Early development of ocular dominance columns. *Science, 290*, 1321–1324.

Curtiss, S. G. (1977). *A psycholinguistic study of a modern-day "wild child."* New York: Academic Press.

Cynader, M., & Mitchell, D. E. (1980). Prolonged sensitivity to monocular deprivation in dark-reared cats. *Journal of Neurophysiology, 43*, 1026–1040.

Daw, N. W. (1997). Critical periods and strabismus: What questions remain? *Optometry and Vision Science, 74*, 690–694.

DeBello, W. M., & Knudsen, E. I. (2004). Multiple sites of adaptive plasticity in the owl's auditory localization pathway. *Journal of Neuroscience, 24*, 6853–6861.

Ellis, A. W., & Lambon, R. M. A. (2000). Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: Insights from connectionist networks. *Journal of Experimental Psychology: Learning, Memory & Cognition, 26*, 1103–1123.

Fagiolini, M., Fritschy, J.-M., Löw, K., Möhler, H., Rudolph, U., & Hensch, T. K. (2004). Specific GABA circuits for visual cortical plasticity. *Science, 303*, 1681–1683.

Farran, D. C. (2000). Another decade of interventions for children who are low income or disabled: What do we know now? In J. P. Shonkoff & S. J. Meisels (Eds.), *Handbook of early childhood intervention* (2nd ed., pp. 510–548). New York: Cambridge University Press.

Feldman, D. E., & Knudsen, E. I. (1998). Experience-dependent plasticity and the maturation of glutamatergic synapses [Review]. *Neuron, 20*, 1067–1071.

Flavell, S. W., Cowan, C. W., Kim, T.-K., Greer, P. L., Lin, X., Paradis, S., et al. (2006). Activity-dependent regulation of MEF2 transcription factors suppresses excitatory synapse number. *Science, 311*, 1008–1012.

Flege, J. E., MacKay, I. R., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of Acoustical Society of America, 106*, 2973–2987.

Flege, J. E., Munro, M. J., & MacKay, I. R. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America, 97*, 3125–3134.

Frieda, E. M., Walley, A. C., Flege, J. E., & Sloane, M. E. (1999). Adults' perception of native and nonnative vowels: Implications for the perceptual magnet effect. *Perception & Psychophysics, 61*, 561–577.

Friederici, A. D. (2006). The neural basis of language development and its impairment. *Neuron, 52*, 941–952.

Fukuchi-Shimgori, T., & Grove, E. A. (2001). Neocortex patterning by the secreted signaling molecule FGF8. *Science, 294*, 1071–1074.

Fukuchi-Shimgori, T., & Grove, E. A. (2003). Emx2 patterns the neocortex by regulating FGF positional signaling. *Nature Neuroscience, 6*, 825–831.

Ghera, M., Marshall, P., Fox, N., Zeanah, C., Nelson, C. A., & Smyke, A. (2009). Social deprivation and young institutionalized children's attention and expression of positive affect: Effects of a foster care intervention. *Journal of Child Psychology and Psychiatry, 50*, 253–256.

Gordon, N. (2000). The acquisition of a second language [Review]. *European Journal of Pediatric Neurology, 4*, 3–7.

Greenough, W. T., Black, J. E., & Wallace, C. S. (1987). Experience and brain development. *Child Development, 58*, 539–559.

Grossman, A. W., Churchill, J., McKinney, B. C., Kodish, I. M., Otte, S. L., & Greenough, W. T. (2003). Experience effects on brain development: Possible contributions to psychopathology. *Journal of Child Psychology and Psychiatry, 44*, 33–63.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Downs-Pruitt, J. C. (1998). Acquisition of English phonetic categories by Japanese speakers: Evidence from categorical discrimination tests. *Journal of the Acoustical Society of America, 103*, 3089–3090.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America, 107*, 2711–2724.

Hammock, E. A. D., & Levitt, P. (2006). The discipline of neurobehavioral development: The emerging interface that builds processes and skills. *Human Development, 49*, 294–309.

Hanm, Y. K., Köver, H., Insanally, M. N., Semerdjian, J. H., & Bao, S. (2007). Early experience impairs perceptual discrimination. *Nature Neuroscience, 10*, 1191–1197.

He, H.-Y., Ray, B., Dennis, K., & Quinlan, E. M. (2007). Experience-dependent recovery of vision following chronic deprivation amblyopia. *Nature Neuroscience, 10*, 1134–1136.

Hebért, J. M., & Fishell, G. (2008). The genetics of early telecephalon patterning: Some assembly required. *Nature Reviews Neuroscience, 9*, 678–685.

Hensch, T. K. (2005). Critical period mechanisms in developing visual cortex. *Current Topics in Developmental Biology, 69*, 215–237.

Hensch, T. K., & Stryker, M. P. (2004). Columnar architecture sculpted by GABA circuits in developing cat visual cortex. *Science, 303*, 1678–1681.

Hess, E. H. (1973). *Imprinting: Early experience and the developmental psychobiology of attachment.* New York: Van Nostrand Reinhold.

Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development, 73*, 418–433.

Horn, G. (2004). Pathways of the past: The imprint of memory. *Nature Reviews Neuroscience, 5*, 108–120.

Huttenlocher, P. R., & Dabholkar, A. S. (1997). Regional differences in synaptogenesis in human cerebral cortex. *Journal of Comparative Neurology, 387*, 167–178.

Itard, J. M. G. (1932). *The wild boy of Aveyron* (G. Humphrey & M. Humphrey, Trans. 1967). New York: Century. [Original work published 1801]

Jeffaris, H., Power, C., & Hertzman, C. (2002). Birth weight, childhood socioeconomic environment and cognitive development in the 1958 British birth cohort study. *British Medical Journal, 325*, 305–311.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*, 60–99.

Jones, E. G. (2000). Cortical and subcortical contributions to activity-dependent plasticity in primate somatosensory cortex. *Annual Review of Neuroscience, 23*, 1–37.

Karmarkar, U. R., & Dan, Y. (2006). Experience-dependent plasticity in adult visual cortex. *Neuron, 52*, 577–585.

Katz, L. C., & Shatz, C. J. (1996). Synaptic activity and the construction of cortical circuits. *Science, 274*, 1133–1138.

Keuroghlian, A. S., & Knudsen, E. I. (2007). Adaptive auditory plasticity in developing and adult animals. *Progress in Neurobiology, 82*, 109–121.

Kim, D. S., & Bonhoeffer, T. (1994). Reverse occlusion leads to precise restoration of orientation preference maps in visual cortex. *Nature, 370*, 370–372.

Knudsen, E. I. (2004). Sensitive periods in the development of the brain and behavior. *Journal of Cognitive Neuroscience, 16*, 1412–1425.

Komarova, N. L., & Nowak, M. A. (2001). Natural selection of the critical period for language acquisition. *Proceedings of the Royal Society of London. Series B, Biological Sciences, 268*, 1189–1196.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience, 5*, 831–843.

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2001). Early visual experience and face processing. *Nature, 410*, 890.

Levitt, P. (2000). Development of the vertebrate nervous system. In M. Wong-Riley (Ed.), *Neuroscience secrets* (pp. 46–57). Philadelphia: Hanley & Belfus.

Levitt, P. (2003). Structural and functional maturation of the developing primate brain. *Journal of Pediatrics, 143*, S35–S45.

Long, M. H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition, 12*, 251–285.

Majdan, M., & Shatz, C. J. (2006). Effects of visual experience on activity-dependent gene regulation in cortex. *Nature Neuroscience, 9*, 650–659.

Marshall, P., Reeb, B. C., Fox, N. A., & the BEIP Core Group (2008). Effects of early intervention on EEG power and coherence in previously institutionalized children in Romania. *Development and Psychopathology, 20*, 845–859.

Martin, S. L., Ramey, C. T., & Ramey, S. L. (1990). The prevention of intellectual impairment in children of impoverished families: Findings of a randomized trial of educational daycare. *American Journal of Public Health, 80*, 844–847.

Mataga, N., Mizuguchi, Y., & Hensch, T. K. (2004). Experience-dependant pruning of dendritic spines at visual cortex by tissue plasminogen activator. *Neuron, 44*, 1031–1041.

Maya Vetencourt, J. J., Sale, A., Viegi, A., Baroncelli, L., De Pasquale, R., O'Leary, O. F., et al. (2008). The antidepressant fluoxetine restores plasticity in the adult visual cortex. *Science, 320*, 385–388.

Merabet, L. B., Rizzo, J. F., Amedi, A., Somers, D. C., & Pascual-Leone, A. (2005). What blindness can tell us about seeing again: Merging neuroplasticity and neuroprostheses [Review]. *Nature Reviews Neuroscience, 6*, 71–77.

Miller, K. D., Keller, J. B., & Stryker, M. P. (1989). Ocular dominance column development: Analysis and simulation. *Science, 245*, 605–615.

Morton, J. B., & Munakata, Y. (2005). What's the difference? Contrasting modular and neural network approaches to understanding developmental variability. *Journal of Development and Behavioral Pediatrics, 26*, 128–139.

Munakata, Y., & Pfaffly, J. (2004). Hebbian learning and development. *Developmental Science, 7*, 141–148.

Nelson, C. A., de Haan, M., & Thomas, K. M. (2006). Neural bases of cognitive development. In W. Damon, R. Lerner, D. Kuhn, & R. Siegler (Vol. Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (6th. ed., pp. 3–57). New York: Wiley.

Nelson, C. A., & Jeste, S. (2008). Neurobiological perspectives on developmental psychopathology. In M. Rutter, D. Bishop, D. Pine, S. Scott, J. Stevenson, & E. Taylor, et al. (Eds.), *Textbook on child and adolescent psychiatry* (5th ed., pp. 145–159). Oxford, UK: Blackwell.

Nelson, C. A., Zeanah, C. H., Fox, N. A., Marshall, P. J., Smyke, A., & Guthrie, D. (2007). Cognitive recovery in socially deprived young children: The Bucharest Early Intervention Project. *Science, 318*, 1937–1940.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science, 14*, 11–28.

Newport, E. L., Bavelier, D., & Neville, H. J. (2001). Critical thinking about critical periods: Perspectives on a critical period for language acquisition. In E. Doupoux (Ed.), *Language, brain and cognitive development: Essays in honor of Jacques Mehler* (pp. 481–502). Cambridge, MA: MIT Press.

O'Leary, D. D. M., Chou, S. J., & Sahara, S. (2007). Area patterning of the mammalian cortex. *Neuron, 56*, 252–269.

Oray, S., Majewska, A., & Sur, M. (2004). Dendritic spine dynamics are regulated by monocular deprivation and extracellular matrix degradation. *Neuron, 44*, 1021–1030.

Oyama, S. (1990). The idea of innateness: Effects on language and communication research. *Developmental Psychobiology, 23*, 741–747.

Paradis, S., Harrar, D. B., Lin, Y., Koon, A. C., Hauser, J. L., Griffith, E. C., et al. (2007). An RNAi-based approach identifies molecules required for glutamatergic and GABAergic synapse development. *Neuron, 53*, 217–232.

Pascalis, O., de Haan, M., & Nelson, C. A. (2002). Is face processing species specific during the first year of life? *Science, 296*, 1321–1323.

Pascalis, O., Scott, L. S., Kelly, D. J., Dufour, R. W., Shannon, R. W., Nicholson, E., et al. (2005). Plasticity of face processing in infancy. *Proceedings of the National Academy of Sciences of the United States of America, 102*, 5297–5300.

Pascual-Leone, A., Amedi, A., Fregni, F., & Merabet, L. B. (2005). The plastic human brain cortex. *Annual Review of Neuroscience, 28*, 377–401.

Puelles, L. (2001). Evolution of the nervous system: Brain segmentation and forebrain development in amniotes. *Brain Research Bulletin, 55*, 695–710.

Rakic, P. (2005). Less is more: Progenitor cell death and cortical size. *Nature Neuroscience, 8*, 981–982.

Rakic, P. (2006). A century of progress in corticoneurogenesis: From silver impregnation to genetic engineering. *Cerebral Cortex, 16*(Suppl.), 1–17.

Sabatini, M. J., Ebert, P., Lewis, D. A., Levitt, P., Cameron, J. L., & Mirnics, K. (2007). Amygdala gene expression correlates of social behavior in monkeys experiencing maternal separation. *Journal of Neuroscience, 27*, 3295–3304.

Sánchez, M. M., Ladd, C. O., & Plotsky, P. M. (2001). Early adverse experience as a developmental risk factor for later psychopathology: Evidence from rodent and primate models. *Developmental Psychopathology, 13*, 419–449.

Scarr, S., & McCartney, K. (1983). How people make their own environments: A theory of genotype environment effects. *Child Development, 54*(2), 424–435.

Scherf, K. S., Behrmann, M., Humphreys, K., & Luna, B. (2007). Visual category-selectivity for faces, places and objects emerges along different developmental trajectories. *Developmental Science, 10*, F15–F30.

Seidenberg, M. S., & Zevin, J. D. (2006). Connectionist models in developmental cognitive neuroscience: Critical periods and the paradox of success. In Y. Munakata & M. Johnson (Eds.), *Processes of change in brain and* *cognitive development. Attention and performance XXI* (pp. 585–612). Oxford, UK: Oxford University Press.

Sheng, M., & Hoogenraad, C. C. (2007). The postsynaptic architecture of excitatory synapses: A more quantitative view. *Annual Review of Biochemistry, 76*, 823–847.

Shepherd, J. D., & Huganir, R. L. (2007). The cell biology of synaptic plasticity: AMPA receptor trafficking. *Annual Review of Cell and Developmental Biology, 23*, 613–643.

Singer, W. (1995). Development and plasticity of cortical processing architectures. *Science, 270*, 758–764.

Stanwood, G. D., & Levitt, P. (2008). The effects of monoamines on the developing nervous system. In C. A. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (2nd ed., pp. 83–94). Cambridge, MA: MIT Press.

Stein, M., Dierks, T., Brandeis, D., Wirth, M., Srtik, W., & Koenig, T. (2006). Plasticity in the adult language system: A longitudinal electrophysiological study on second language learning. *Neuroimage, 33*(2), 774–783.

Sudhof, T. C. (2008). Neurotransmitter release. *Handbook of Experimental Pharmacology, 184*, 1–21.

Sugita, Y. (2008). Face perception in monkeys reared with no exposure to faces. *Proceedings of the National Academy of Sciences of the United States of America, 105*, 394–398.

Sugiyama, S., Di Nardo, A. A., Aizawa, S., Matsuo, I., Volovitch, M., Prochiantz, A., et al. (2008). Experience-dependant transfer of Otx2 homeoprotein into the visual cortex activates post natal plasticity. *Cell, 134*, 508–520.

Sur, M., & Rubenstein, J. L. R. (2005). Patterning and plasticity of the cerebral cortex. *Science, 310*, 805–810.

Szyf, M., McGowan, P., & Meany, M. J. (2008). The social environment and the epigenome. *Environmental and Molecular Mutagenesis, 49*, 46–60.

Tomblin, J. B., Barker, B. A., Spencer, L., Zhang, X., & Gantz, B. J. (2005). The effect of age at cochlear implant initial stimulation on expressive language growth in infants and toddlers. *Journal of Speech, Language and Hearing Research, 48*, 853–867.

Trachtenberg, J. T., & Stryker, M. P. (2001). Rapid anatomical plasticity of horizontal connections in the developing visual cortex. *Journal of Neuroscience, 15*, 3476–3482.

Tritsch, N. X., Yi, E., Gale, J. E., Glowatski, E., & Bergles, D. E. (2007). The origin of spontaneous activity in the developing auditory system. *Nature, 450*, 50–55.

von Melchner, L., Pallas, S. L., & Sur, M. (2000). Visual behaviour mediated by retinal projections directed to the auditory pathway. *Nature, 404*, 871.

Wartenburger, I., Heekeren, H. R., Abutalebi, J., Cappa, S. F., Villringer, A., & Perani, D. (2003). Early setting of grammatical processing in the bilingual brain. *Neuron, 9*, 159–170.

Weber-Fox, C., & Neville, H. J. (2001). Sensitive periods differentiate processing of open- and closed-class words: An ERP study of bilinguals. *Journal of Speech, Language, and Hearing Research, 44*, 1338–1353.

Werker, J. F., & Tees, R. C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology, 46*, 233–251.

Wiesel, T. N., & Hubel, D. H. (1963). Single-cell responses in striate cortex of kittens deprived of vision in one eye. *Journal of Neurophysiology, 26*, 1003–1017.

Wiesel, T. N., & Hubel, D. H. (1965). Extent of recovery from the effects of visual deprivation in kittens. *Journal of Neurophysiology, 28*, 1060–1072.

Windsor, J., Glaze, L. E., Koga, S. F., & the BEIP Core Group. (2007). Language acquisition with limited input: Romanian institution and foster care. *Journal of Speech, Language, and Hearing Research, 50*, 1365–1381.

Wonders, C. P., & & Anderson, S. A. (2007). The origins and specification of cortical interneurons. *Nature Reviews Neuroscience, 7*, 687–696.

Zingg, R. M. (1940). Feral man and extreme cases of isolation. *The American Journal of Psychology, 53*, 487–517.

Child Development, January/February 2010, Volume 81, Number 1, Pages 41–79

Epigenetics and the Biological Definition of Gene × Environment Interactions

Michael J. Meaney
McGill University

Variations in phenotype reflect the influence of environmental conditions during development on cellular functions, including that of the genome. The recent integration of epigenetics into developmental psychobiology illustrates the processes by which environmental conditions in early life structurally alter DNA, providing a physical basis for the influence of the perinatal environmental signals on phenotype over the life of the individual. This review focuses on the enduring effects of naturally occurring variations in maternal care on gene expression and phenotype to provide an example of environmentally driven plasticity at the level of the DNA, revealing the interdependence of gene and environmental in the regulation of phenotype.

The nature–nurture debate is essentially a question of the determinants of individual differences in the expression of specific traits among members of the same species. The origin of the terms *nature* and *nurture* has been credited to Richard Mulcaster, a British teacher who imagined these influences as collaborative forces that shape child development (West & King, 1987). History has conspired to pervert Mulcaster's intent, casting genetic and environmental influences as independent agents in the field of development. The intent of this article is to examine the biological context in which genetic and environmental signals actually operate. The result of this analysis produces an impression that lies close to Mulcaster's formulation.

Any successful attempt to constructively leverage the remarkable advances of the genomic era will depend upon our ability to understand genetic influences and their interactions with the environmental context within which they operate. Hindering such efforts is the rather arcane notion that we can partition the causes of individual differences into distinct genetic *or* environmental spheres of influence. This issue is of particular importance in the area of child development where research that examines the origins of individual differences in complex traits often forms the basis for extensive intervention or prevention programs (Fisher, Gunnar, Chamberlain, & Reid, 2000; Olds et al., 1998). Our assumptions concerning the processes of child development influence the design of such programs. Thus, one of the great challenges at this time is that of integrating genomics into the design of prevention and intervention programs. In the social sciences, and particularly psychology, there has generally been an understandable bias in favor of explanations derived from the "nurture" perspective, which emphasizes the capacity for environmentally induced plasticity in brain structure and function. Nevertheless, over recent decades there has been a gradual, if at times reluctant acceptance that genomic variations contribute to individual differences in brain development and function. And thus the integration of genetics into psychology and psychiatry has become one of the more active research areas in the study of neural function. Now that we have come to accept the idea that genomic variation influences brain development and function in humans, it is time to consider what this should mean in the reformulation of intervention programs that target children and their families.

The idea that variation in the expression of neural functions would lie outside the dominion of genomic influences is biologically untenable. Like any tissue, the development of the mammalian brain and its function occurs as a result of coordinated influences that include heritable variation in genomic sequence. What remains as a major challenge for the psychological sciences is the development of a conceptual framework from which to meaningfully understand the results of studies that examine brain development and function at the level of the genotype. However, we remain mired in additive models of explanation within which phenotypic development rolls out as the cumulative by product of genetic + environmental

---

Correspondence concerning this article should be addressed to Michael J. Meaney, Douglas Mental Health University Institute, McGill University, 6875 LaSalle Blvd, Montreal, QC, Canada H4H 1R3. Electronic mail may be sent to michael.meaney@mcgill.ca.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8101-0005

influences, with each discipline adding independently to the equation. Science in the postgenomic era requires a more sophisticated, biologically informed understanding of the interplay between gene and environment in defining individual differences at the level of complex traits.

The fundamental question is that of the relation between genotype and phenotype: How do variations in the genotype influence selected traits at any point over the lifespan? This question is addressed very differently in epidemiology and the social sciences than it is in the biological sciences. Indeed, the actual meaning and interpretation of the term *gene* can differ across research domains (Griffiths & Tabery, 2008). For epidemiologists or those employing epidemiological approaches, genotype–phenotype relations are defined by statistical probabilities. A gene *for* aggression is thus defined as a genomic region within which a defined variation in sequence predicts a specific phenotypic outcome, in this case one associated with aggressive thoughts or actions. Here, genotype–phenotype relations are statistical. For the biological scientist, genotype–phenotype relations are defined by the actual physical operation of a genomic region in relation to the proximal cellular events that directly mediate the behavioral variation as well as the more distal influences that regulate the relevant cellular signals. Thus, a gene *for* aggression would be defined as a genomic region from which a transcribed product actively participates in a biological pathway that *directly* mediates the expression of aggressive behaviors. The studies of Young and colleagues (Donaldson & Young, 2008; Hammock & Young, 2005) on the sequence variation in the vasopressin 1A receptor gene provide an excellent example. Among voles, subspecies differences in social behaviors are associated with a sequence variation in the vasopressin 1A receptor gene and differential expression of the vasopressin 1A receptor. Reversing the pattern of vasopressin 1A receptor expression eliminates the differences in selected, vasopressin-mediated social behaviors (see Robinson, Fernald, & Clayton, 2008, for additional examples). These studies describe a causal genotype–phenotype relation. Within the biological sciences, the mere statistical relation between a genomic variation and phenotypic outcome would not be considered as sufficient basis for the establishment of a functional link between the gene (or its product) and the relevant variation in phenotype. Such differences in perspective have led to rather lively debates between, for example, researchers in quantitative behavioral genetics and researchers examining the biological basis for individual differences in neural function (Griffiths & Stotz, 2007; Griffiths & Tabery, 2008; Moss, 2005).

The issue becomes even more complicated when considering the interaction between gene and environment, since the statistical approaches to the study of Gene × Environment interactions are complicated and not yet fully developed. Indeed, since the time of Fisher and Haldane there has been debate concerning the adequacy of analysis of variance (ANOVA) models in detecting Gene × Environment interactions (Wahlsten, 1990; see also Griffiths & Stotz, 2007; Griffiths & Tabery, 2008, for historical reviews). The primary objective of this article is to examine the Gene × Environment interaction from the biological perspective. For whatever the difficulties in statistically defining Gene × Environment interactions, research in biology reveals that the genome cannot possibly operate independent of its environmental context. The biological perspective reveals the futility of the nature–nurture debate and of additive models of genetic and environmental influences in defining phenotype. Moreover, recent studies from the field of epigenetics provide insight into biological mechanisms whereby Gene × Environment interactions can biochemically alter the genome and thus stably influence individual differences in neural function. Developmentalists studying the enduring effects of early experience on brain development and function have long anticipated such processes (Bateson et al., 2004; Gottlieb, 1997, 1998; Schneirla, 1966). This approach envisions development as an active process of adaptation that occurs as a function of the continuous dialog between the genome and its environment. With the integration of epigenetics into developmental psychobiology, we can now begin to define the physical basis for this interactive process.

Statistical Approaches to the Study of the "Instrumental" Gene

Research examining the potential influence of heritable variations in DNA sequences on brain development and function in humans has relied largely upon two approaches. The first is a classical epidemiological approach that involves association or linkage analyses that essentially use correlational approaches to describe genotype–phenotype relations. Much as epidemiological research statistically associates specific exposures with health outcomes, such studies correlate variation in nucleotide sequence at specific sites in the genome with a phenotypic outcome. This is certainly a reasonable approach in the initial phase of examining the *potential* role for a specific genomic variant (Plomin & Rutter, 1998). The advent of high-throughput technology permits genome-wide association studies that can potentially scan 1 million or more single nucleotide polymorphisms in individual subjects. While there are inherent statistical complications, such as controlling for multiple comparisons, the notion of being able to examine the relation between 10^6 sequence variants with a single trait of interest across a sufficiently large sample is incredibly seductive. Nevertheless, such studies do not inform at the level of causal mechanism. A correlation between a genomic variant and a developmental outcome is no less correlational and no more explanatory (or mechanistic) than is the correlation between, say, level of education or a dietary constituent and the same developmental outcome. The popular press often trumpets such findings (usually simply echoing the institute's own press release) as the identification of a gene *for* an outcome of interest. Studies that, for example, statistically identify a relation between a genomic polymorphism and the variation in the occurrence of conduct disorder are interpreted as successfully revealing the gene *for* conduct disorder. Such interpretations drive many scientists slightly mad, since the data are merely correlational and do not imply a cause–effect relation between the genomic variant and conduct disorder. The gene might indeed be expressed only in the liver and influence the development or activity of the relevant neural systems that regulate social behavior only through a very general metabolic pathway that fuels brain development in early life. There may be no direct effect on any neuronal population whatsoever, let alone the complex neural network that underlies the ability to regulate affect and behavior. Nevertheless, the genomic variant is indeed statistically associated with conduct disorder. The controversy surrounding the claims from association or linkage studies lie in the interpretation of the data and the implicit confusion of correlation with causation. The data are safely assumed to provide the identification of *potential* influence over the phenotype in question. However annoying the misuse of such correlational data in public or professional discourse, the findings serve as a starting point for research examining the possible functional relation between the genomic variant and the phenotype. In the case of our example, the nature of the relation between the genomic variant of interest and conduct disorder must then become a matter of more precise studies that directly examine the functional significance of that genomic variant, as well as the particular circumstances under which the genotype–phenotype relation is apparent (Plomin & Rutter, 1998).

A second statistical approach is one that has evoked greater controversy. Quantitative behavioral genetics does not examine genotype–phenotype relations at the level of specific genomic variants but rather attempts to partition the variance in the expression of a specific phenotypic traits across a population into that associated with "genetic" or "environmental" influences (Plomin, DeFries, & McClearn, 1990). Such studies actually examine so-called genetic effects without ever directly examining the genome let alone variation in the genome (Sokolowski & Wahlsten, 2001). Indeed within quantitative behavioral genetics the issue that is directly under study is that of heritability. Genetic effects are simply implied on the basis of data derived from studies of the heritability of traits. This statistical approach, based on the development of ANOVA models by Fisher and Haldane at the turn of the last century, has a long history in agriculture and animal sciences as the basis for selective breeding for specific traits (Griffiths & Tabery, 2008; Wahlsten, 2003). Such endeavors do not rely on the assumption that heritability necessarily reflects a genomic mechanism for the transmission of individual differences in the trait from parent to offspring. What is critical for selective breeding is the ability to reliably estimate the passage of traits from the parent to offspring. Traits with greater heritability are better candidates for selective breeding. The actual mechanism of inheritance is irrelevant. I think that the situation in the biological sciences is mixed. For some, heritability simply reflects familial transmission. For others, it has come to be equated with a true estimate of variation due to variation in genomic sequence. Estimates of heritability are commonly equated with "genetic" influences. It is this assumption that proves controversial.

Johannsen (1911) first introduced the terms *genotype* and *phenotype* and the relation between the two became perhaps the defining question in biology. In the study of genotype–phenotype relations, the term *phenotype* refers to all aspects of gene function and may thus be used to describe any observable characteristic of an organism, such as its morphology, biochemical or physiological properties, or behavior (Johannsen, 1911; Lewontin, 1974). The gene had actually been the subject of scientific analysis well before the early 20th century and long before the identification of DNA as the physical basis for the "genetic code." Over this period, the gene was considered as the hypothetical physical constituent of the cell, the transmission of which from parent to offspring served as the basis for Mendelian inheritance (Griffiths & Stotz, 2007; Griffiths & Tabery, 2008). The gene was defined as the "unit of heredity." And thus it was understandable that during this period estimates of heredity were assumed to reflect a genetic influence. This was true by definition. Griffiths and Tabery (2008) thus refer to this entity as the "traditional gene" —the substance of inheritance. For the early pioneers in genetics such as Morgan, and many studying patterns of inheritance to this day, the actual physical identity and operation of the gene is irrelevant (Falk, 1986). This tradition continues within the study of quantitative behavioral genetics for which the gene is simply viewed as the unit of inheritance, with apparently little concern for the actual biology that might mediate the relation of the genomic variant, cellular function and the trait of interest.

A classical research approach in quantitative behavioral genetics has employed studies of twins to infer "genetic" influences. Twin studies rely on the level of similarity in specific traits between monozygotic and dizygotic twins to derive estimates of heritability (h^2). As noted earlier, such measures were developed in animal sciences using parent–offspring regressions to study intergenerational transmission of specific traits. The intent of such studies is not to identify the causal interplay between gene and environment in defining the development and expression of a particular trait in an individual but rather to estimate the relative contribution of genetic and environmental influences in defining the variation in the trait across a population (Plomin et al., 1990). The focus on "relative contributions" of gene and environment as *independent* forces is the hallmark of quantitative behavioral genetics and assumes that such forces can act independently of one another. In following the biometric tradition of Fisher and Haldane, the variance in the expression of a phenotypic trait across a population is examined using ANOVA models that partition variance in phenotype according to genetic and environmental factors. Note that within such analyses the "genetic" influences are defined *statistically* and not biologically, and of course the same is true for the definition of so-called environmental influences. Moreover, the resulting analyses are an attempt to understand the source of variation in phenotype across a population and not as an explanation for individual patterns of development (Griffiths & Tabery, 2008).

Studies of the heritability of selected traits thus yield *statistical* and not *biological* evidence for a genetic influence. The irritation derives from the interpretation of such findings, and in particular the notion of "main effects," which implies that the genome and its environment can operate independently. And of course the angst should be no less for environmental main effects than for those associated with genotype. Again the issue is that of interpretation and an appreciation of what such findings actually do and do not represent.

Twin studies derive estimates of heritability. Such findings are of considerable importance for their own sake, particularly for those interested in issues surrounding familial transmission of vulnerability for illness. An indication of high heritability for a specific trait logically leads to the question of whether genomic variations at the level of sequence might serve as the mechanism for the passage of individual differences in phenotype from parent to offspring (prior to investing valuable research funds in genome-wide searchers, it is indeed useful to establish that the variation in the trait is actually subject to familial transmission). But the idea that estimates of heritability are the *equivalent* of a genetic effect is problematic. First, such estimates ignore Gene × Environment interactions. Indeed, heritability equates with a main effect "genetic" influences only if one assumes no Gene × Environment contribution. Second, the estimate of heritability reflects the importance of genomic mechanisms only if one assumes that there are no other biological mechanisms for inheritance. This is simply untrue. There are multiple potential mechanisms of inheritance, involving, for example, the passage of epigenetic marks (Chong & Whitelaw, 2004) through the germline, the passage of maternal RNA molecules into the embryo, the potential passage of prion proteins from parent to offspring, the biochemical state of the gametes at the time of conception, and the transmission of nutrients, bacteria, or antibodies from maternal circulation to that of the offspring, and so on (Bettegowda & Smith, 2007; Boulinier & Staszewski, 2008; Chong & Whitelaw, 2004; Grindstaff, Brodie, & Ketterson, 2003; Hasselquist & Nilsson, 2009; Rassoulzadegan et al., 2006; Shorter & Lindquist, 2005). Variations in genomic sequence are simply not the only mechanism of inheritance. All of these factors can and do influence the phenotype of the offspring. This is a serious problem for those assuming that heritability equates with main effects of genomic variation.

Apart from the methodological and biological pitfalls with such inferences (Lewontin, 1974; Sokolowski & Wahlsten, 2001; Wahlsten, 1990), the argument presented here is that the notion of a main effect of either ''gene'' or ''environment'' on the development of individual differences in complex traits is biologically fallacious. Instead, the development of the individual is best considered as the emergent property of a constant interplay between the genome and its environment (Gottlieb, 1991, 1997). This article briefly examines how the study of gene activity or expression reveals why main effects models are *biologically* implausible and then reviews studies on the effects of maternal care on glucocorticoid receptor expression in the rat as a model for the understanding one process by which Gene × Environment interactions might operate to influence the development of individual differences.

The Molecular Gene and Its Operation

The challenge within developmental psychology is that of understanding the operation of the genome and how the analysis of nucleotide-based variation can contribute to our understanding of individual differences in brain development and function at the level of cell biology, physiology, and emotional-cognitive states. One obvious way to start is to simply appreciate the role of DNA in cell biology and realize that DNA *directly* codes for molecules, more specifically for ribonucleic acids (RNAs), and not for higher levels of function. While there are indeed statistical relations between variation in nucleotide sequence and that in complex traits, at the level of biology there are no genes for intelligence, depression, athletic abilities, fashion sense, or any other such complex trait. Rather, there are certain variations in genomic sequences that can potentially alter either the DNA product (RNA) or the degree to which the DNA is transcribed (i.e., actively producing its molecular product). There are multiple and complex cellular processes that lie between the DNA sequence and the functional outcome associated with the gene product. The relation between genotype and phenotype, even at the level of cellular molecules, is anything but direct. Thus, the intent of this rather brief section is to simply illustrate the complexity of the processes that mediate genotype–phenotype relations even when considering the most fundamental aspects of gene function—the production of RNAs and protein products.

Throughout portions of this review I will largely focus on one gene, that encoding for the glucocorticoid receptor, both as an exemplar, for which it is rather well suited, and as a subject for our own research. The glucocorticoid receptor belongs to the nuclear receptor superfamily and is transcribed from a single mRNA. The glucocorticoid receptor protein is found both in the cytoplasm and in the nucleus of the cell. The receptor belongs to a class of intracellular proteins referred to as ligand-activated transcription factors. This simply refers to the fact that the binding of the hormone (the ligand) to the receptor alters the configuration of the glucocorticoid receptor, which then permits binding to specific sites on the DNA and the regulation of gene transcription. Glucocorticoids, a hallmark of the endocrine stress response, are hormones secreted by the adrenal cortex that bind to the glucocorticoid receptor to mediate various cellular processes including those involved in cardiovascular activity, appetite, metabolism, immune responses, electrolyte balance, as well as neuronal function and behavior (e.g., Dallman, Akana, Strack, Hanson, & Sebastian, 1995; Dallman, Pecoraro, & la Fleur, 2005; de Kloet, Karst, & Joels, 2008; Heitzer, Wolf, Sanchez, Witchel, & DeFranco, 2007; Kumar & Thompson, 2005; Munck, 2005; Munck, Guyre, & Holbrook, 1984; Sapolsky, Romero, & Munck, 2000; Seckl & Holmes, 2007; Walker, 2007). Thus, the gene encoding the glucocorticoid receptor has multiple, tissue-specific functions. The activation of the glucocorticoid receptor alters gene expression, the number and identity of which depends very much upon the target cell. When not bound by glucocorticoids, the glucocorticoid receptor resides in the cytoplasm of the cell associated with what is termed a chaperone protein (Kumar & Thompson, 2005; Lu & Cidlowski, 2006; McNally, Müller, Walker, Wolford, & Hager, 2000). Binding of glucocorticoids induces conformational changes in the receptor; dissociation from the chaperone proteins; association with another, activated glucocorticoid receptor protein, thus forming a dimer (association of two proteins); translocation into the nuclear compartment; and binding to specific sites on the DNA (Lu & Cidlowski, 2006 and references therein). The activated glucocorticoid receptor can also complex with other proteins, or cofactors, that then modify the effect of glucocorticoid receptor on the expression of target genes. The association of the cofactor may simply modify the ability of the glucocorticoid receptor to access DNA sites, and thus the magnitude of its transcriptional effect. However, the presence of the cofactor may determine the nature of the affected genomic target. Thus, an activated glucocorticoid receptor associated with factor A may influence a different set of genes than the activated glucocorticoid receptor associated with factor B. Thus, the glucocorticoid receptor can affect two different cellular populations in very different ways depending upon the associated cofactor. The function of the glucocorticoid receptor gene product depends upon the context within which it operates. Indeed, the presence of other transcriptional factors can actually determine the direction of the effect of the glucocorticoid receptor on a common genomic target, increasing transcription in one context and decreasing it in another (e.g., Diamond, Miner, Yoshinaga, & Yamamoto, 1990). For example, activation of the glucocorticoid receptor promotes cell survival in liver cells but activates apoptosis (cell death) in thymocytes. Thus, a sequence-based variation in the glucocorticoid receptor could (and indeed does) have exactly the opposite effect on two different cellular populations. Of course, not all contextual effects are quite so dramatic. But the influence of such cofactors reveals a complexity that suggests that the consequences of even a highly functional sequence variation in the glucocorticoid receptor gene, as an example, will be determined by the cellular mileu in which the glucocorticoid receptor gene is operational. Some cells would inevitably be more affected than others. The presence and activity of such cofactors is regulated by a variety of extra- and intracellular signals and forms an important feature of the context within which glucocorticoid receptor activity is defined. This rather complicates the study of genotype–phenotype relations and the plausibility of simple main effects of either genome or environment at the level of cell biology. Rather, such processes suggest that the effect of genomic sequence variation is understood only in terms of its interaction with environmental signals represented in the cellular context. Such signals, of course, would also reflect sequence variation in other genomic regions. Since the transcriptional cofactors that influence glucocorticoid receptor activity are constantly regulated by the internal and external environment, the functional implications of a sequence-based variation in the glucocorticoid receptor gene are inexorably linked to context.

The glucocorticoid receptor can also alter gene expression and cellular function through processes that do not involve the interaction of the receptor with DNA. Intracellular proteins interact to form complexes. In certain cases, such interactions inhibit the action of one of the partners. One of the primary biological effects of the activated glucocorticoid receptor is to dampen inflammation (Munck, 2005; Munck et al., 1984). This effect involves, in part, the interaction of the glucocorticoid receptor with another protein, NFKß (e.g., Nissen & Yamamoto, 2000), which is a major participant in the activation of inflammatory responses. The binding of the glucocorticoid receptor renders NFKß unable to access its DNA targets, thus muting the inflammatory reaction. However, this glucocorticoid receptor effect is apparent only if there is an elevation in NFKß. Thus, the effects of either a genomic or environmental influence would depend upon the immunological status of the organism, further reflecting the importance of context. The function of the glucocorticoid receptor gene with such protein–protein interactions further illustrates why notions of main effects of gene or environment at the level of biological mechanism are untenable.

Figure 1 outlines the organization of the glucocorticoid receptor gene in the rat, which was cloned in large part by the group of Yamamoto (Miesfield et al., 1984) with a recent addition from our group in collaboration that of Jonathan Seckl (McCormick et al., 2000). The organization of the gene is similar in rat and human (Nobukuni, Smith, Hager, & Detera-Wadleigh, 1995; Turner & Muller, 2005). The glucocorticoid receptor gene comprised nine exons (an exon is a region of DNA that contributes directly to an mRNA molecule). Exons 2–9 are coding regions that produce mRNA for amino acids sequences included in the glucocorticoid receptor protein (as in DNA–RNA–protein). Exon 1 is a

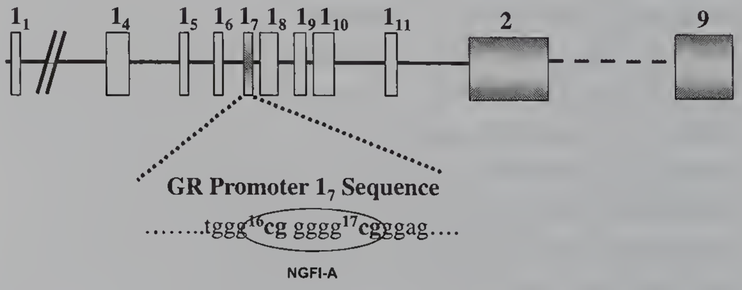

1_1
1_4
1_5
1_6
1_7
1_8
1_9
1_{10}
1_{11}
2
9
GR Promoter 1_7 Sequence
........tggg^{16}cg gggg^{17}cggag....
NGFI-A

Figure 1. A schema describing the organization of the rat glucocorticoid receptor gene including nine exon regions.
Note. Exons 2–9 participate in the coding of the glucocorticoid receptor protein. Exon 1 comprised multiple regions, each of which is capable of activating gene transcription (i.e., promoter sequences). The various exon 1 promoters' actions are tissue-specific, with evidence suggesting that certain promoters are more active in areas such as liver or thymus, and others more active in brain (e.g., exon 17; based on McCormick et al., 2000; see Turner & Muller, 2005, for comparable data in humans). The consensus binding site for NGFI-A lying within the exon 1_7 promoter is highlighted. NGFI-A = nerve growth factor-inducible factor A.

large region containing a number of nucleotide sequences that are represented in the glucocorticoid receptor mRNA, but do not code for protein. This exon 1 region contains a number of regulatory elements, the activation of which alters the level of glucocorticoid receptor gene transcription. This exon 1 region is similarly organized in humans (Turner & Muller, 2005; see Zhang, Haws, & Wu, 2004, for examples of similar genomic organization). In the rat, each of the individual exon 1 segments displayed in Figure 1 contains a region that activates gene transcription (i.e., a transcriptional start site; McCormick et al., 2000) and is therefore referred to as a "promoter." Thus, intracellular signals, referred to as transcription factors, bind to different regions on the exon 1 promoters to activate or repress the transcription of the glucocorticoid receptor gene. Figure 1 shows a small segment from the exon 1_7 promoter for the rat glucocorticoid receptor gene that binds the transcription factor nerve growth factor-inducible factor A (NGFI-A; Crosby, Puetz, Simburger, Fahrner, & Milbrandt, 1991).

The exon 1 region contains a number of promoters, and this complex level of regulation is essential for glucocorticoid receptor function. The glucocorticoid receptor gene is potentially active in virtually every cell in the body, with different functions across various cell types. In the fetal mammal, for example, the activation of the glucocorticoid receptor induces lung surfactant, a protein essential for the transition to a gaseous environment that marks birth. This effect is the basis for the glucocorticoid therapy of premature infants. Increasing glucocorticoid receptor function in pulmonary tissues in late fetal life is essential. Not so in the brain. Activation of glucocorticoid receptors in the mammalian forebrain is associated with a decrease in neurogenesis (Uno, Tarara, Else, Suleman, & Sapolsky, 1989) and synaptic plasticity (de Kloet et al., 2008; McEwen, 2007). Similar catabolic effects are apparent in several other tissues (Munck et al., 1984; Sapolsky et al., 2000). The ideal scenario is that of increasing glucocorticoid receptor expression in the lung while muting that in neurons and other active sites of growth (Bronnegard & Okret, 1991; Sapolsky & Meaney, 1986). Such tissue-specific expression of the glucocorticoid receptor gene is probably accomplished through the activation of different transcriptional promoters in various tissues. Indeed, different promoters are associated with the activation of the glucocorticoid receptor in lung, liver, thymus, and brain (McCormick et al., 2000). It is therefore possible to increase the expression of the glucocorticoid receptor in one tissue while leaving it unaltered or even decreased in another. Thus, a nucleotide-based variant in a glucocorticoid receptor promoter region might lead to altered transcription of the glucocorticoid receptor gene in one tissue, while the activation of the same gene in a different tissue is completely unaffected. The resulting genotype–phenotype interaction would then depend upon whether the trait under study is dependent upon a cellular population that employs the affected promoter sequence.

There is also variation of glucocorticoid receptor gene function at the level of the protein product (Lu & Cidlowski, 2006). Exons 2–9 encode multiple variants of the glucocorticoid receptor. Two of the variants (or isoforms), the α- and β-forms of the glucocorticoid receptor, are identical for the first 727 amino acids, beyond which they differ. Glucocorticoid receptor α has an additional 50 amino acids not contained in the ß-variant. The ß-variant has 15 distinct amino acids. The classic effects of glucocorticoids on gene expression are mediated through glucocorticoid receptor α. Glucocorticoid receptor ß seems to bind to glucocorticoid receptor α and to block its effects on gene transcription (Lu & Cidlowski, 2006). Indeed, increased expression of glucocorticoid receptor ß produces glucocorticoid resistance. What is remarkable is that two seemingly antagonistic proteins are actually encoded from the same gene! Further complicating the picture is the finding that the glucocorticoid receptor ß-variant is produced in only a subset of cells. The production of glucocorticoid receptor variants, and thus of glucocorticoid receptor gene, is entirely dependent upon the cellular context within which the gene operates.

The production of glucocorticoid receptor protein variants is not limited to the α- and ß-isoforms. More recent studies reveal at least eight different versions of the glucocorticoid receptor a protein, each produced from the same gene and even the same mRNA (Lu & Cidlowski, 2006). These glucocorticoid receptor α-isoforms are differentially distributed across various tissues and differ in their transcriptional activity on target genes. Certain genomic targets are regulated by all of the various glucocorticoid receptor α-isoforms, although transcription is regulated to a greater extent by some isoforms than by others. But some genes are regulated only by certain isoforms. In no case is this diversity in the genomic product associated with variation in nucleotide sequence. This complexity is an inherent feature of the glucocorticoid receptor system.

This brief review considers only the variation in the glucocorticoid receptor protein itself. If we factor into consideration the potential for such variation together with the multiple cofactors, each of which can alter glucocorticoid receptor function, the potential for diversity in signaling becomes substantially greater. If genotype–phenotype relations are so remarkably intricate at the level of protein, then imagine the complexity at the level of physiology and behavior! It is critical that we appreciate the degree to which the activity of the gene is dependent upon the cellular context within which it functions and to appreciate that this same context is also subject to the influences of sequence variations in other genes that operate within the same network. Both the form and the function of the protein product are defined by context. The function of the gene can only be fully understood in terms of the cellular environment in which it operates. And the cellular environment, of course, is dynamic, changing constantly as a result of signals from other cells, including those that derive from events occurring in the external environment. Ultimately, function can only be understood in terms of the interaction between environmental signals and the genome.

Gene Transcription

The most compelling evidence for the predominance of Gene × Environment interactions in cellular function emerges from the study of gene transcription. The transcription of a gene is a highly regulated event. At the heart of this process lies a class of proteins referred to as transcription factors (and see earlier). As the name implies, these proteins have the ability to bind to regulatory elements of the gene (e.g., see the exon 1 region of the glucocorticoid receptor gene) and to activate or repress gene transcription. Importantly, the expression and/or activation of the transcription factors themselves are dynamically regulated by environmental signals. Many of the earliest cellular responses to environmental stimuli involve either the activation of preexisting transcriptional signals through, for example, phosphorylation at specific sites, or the rapid synthesis of proteins (e.g., immediate early gene products) that then serve to regulate the activity of other genes. Thus, the binding of transcription factors to DNA sites is the biological machinery of the dynamic Gene × Environment interactions that result in altered rates of gene transcription.

Figure 1 portrays the organization of the glucocorticoid receptor gene. The schema is actually somewhat misleading. For reasons of graphic simplicity, we often describe the organization of a gene or the interactions between transcription factors and DNA as if the DNA were a linear molecule to which transcription factors gain unimpeded access. The reality of protein–DNA interactions is quite different. Figure 2 presents the classic crystallographic analysis of the organization of DNA (Luger, Mader, Richmond, Sargent, & Richmond, 1997). DNA is organized into units referred to as nucleosomes, each of which contains about 145–150 base pairs wrapped around a core region of histone proteins (Turner, 2001). The histones and DNA together are referred to as chromatin; the nucleosome is the organization of chromatin. Under normal conditions there is a tight physical relation between the histone proteins and its accompanying DNA, resulting in a rather closed nucleosome configuration. This restrictive configuration is maintained, in part, by electrostatic bonds between the positively charged histones and the negatively charged DNA. The closed configuration impedes transcription factor binding and is associated with a reduced level of gene expression. The activation of gene expression commonly requires chemical modifica-

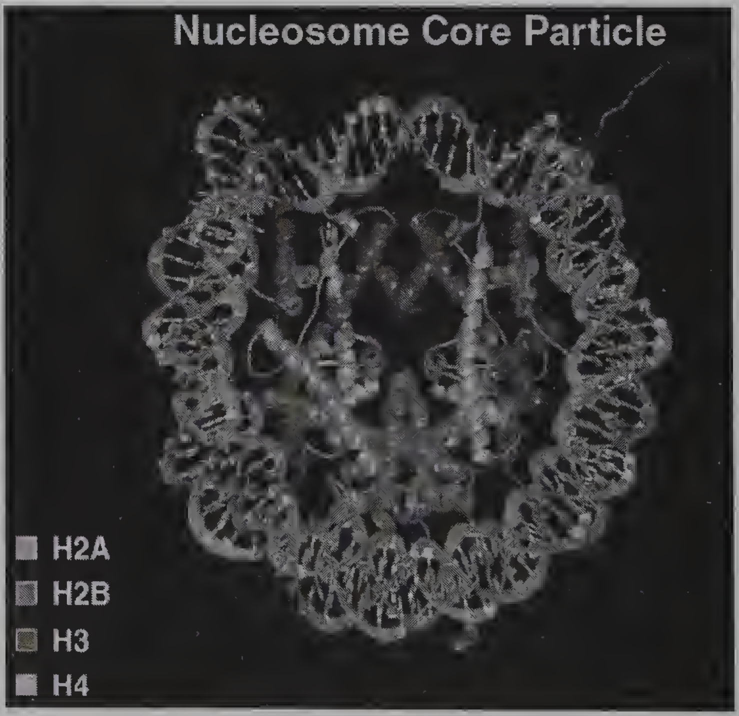


Figure 2. Crystallographic image of the nucleosome showing 146 bp wrapped around a histone complex that comprised histone 2A, 2B, 3, and 4 proteins (from Luger et al., 1997).
Note. The tight configuration is maintained, in part, by electrostatic bonds. Modifications, such as acetylation, to the histone regulate transcription factor binding and occur primarily at the histone tails protruding out of the nuclesome (pictured is the tail of histone 3).

tion of the chromatin that occurs on the histone proteins.

Histone Modifications

Chromatin remodeling is required for increased transcription factor binding to regulatory sites on the DNA and the activation of gene expression. The dynamic alteration of chromatin structure is achieved through modifications to the histone proteins at the "tail" regions that protrude outside of the nucleosome (Figure 2). This process is achieved through a series of enzymes that bind to the histone tails and modify the local chemical properties of specific amino acids along the tail (Grunstein, 1997; Hake & Allis, 2006; Jenuwein & Allis, 2001). The histone tails are chains of amino acids that include lysines and arginines that are primary targets for modification. For example, the enzyme histone acetyltransferase "transfers" an acetyl group onto specific lysines on the histone tails. The addition of the acetyl group diminishes the positive charge, loosening the relation between the histones and DNA, opening the chromatin, and improving the ability of transcription factors to access DNA sites. Thus, histone acetylation at specific lysine sites is commonly associated with active gene transcription. The functional antagonists of the histone acetyltransfereases are a class of enzymes known as histone deacetylases (HDACs). These enzymes remove acetyl groups and prevent further acetylation, thus serving to maintain a closed chromatin structure, decreasing transcription factor and gene expression. The reader should note that there are actually multiple modifications to histone tails including methylation (in this case on the histones), phosphorylation, ubiquitination, and so on. For the sake on simplicity, the discussion is limited to histone acetylation or deacetylation.

Regulation of Glucocorticoid Receptor Expression

Figure 3 summarizes the effect of the neurotransmitter serotonin (5-hydroxytryptamine [5-HT]) on glucocorticoid receptor gene transcription in hippocampal neurons (Mitchell, Betito, Rowe, Boksa, & Meaney, 1992; Mitchell, Rowe, Boksa, & Meaney, 1990; Weaver et al., 2007). This effect is dependent upon the binding of the transcription factor NGFI-A to a specific binding site on the exon 1_7 GR promoter. We can precisely define the importance of this interaction. For example, mutating a single nucleotide, in this case simply exchanging a cytosine for an adenine, in the region of the promoter that normally binds NGFI-A abolishes the ability of NGFI-A to associate with the exon 1_7 promoter and eliminates the effect of NGFI-A on gene transcription (Weaver et al., 2007). However, the ability of NGFI-A to bind to the exon 1_7 promoter is regulated by another protein, a cofactor, the concentration of adenosine 3,5-monophosphate (cAMP)-response element-binding (CREB) protein, that is, activated by the same 5-HT–cAMP/PKA (protein kinase A) signaling cascade (Figure 3). The CREB-binding protein is a histone acetyltransferase. The resulting increase in the association of the CREB-binding protein with the exon 1_7 promoter is accompanied by an increase in the acetylation of a specific lysine on the tail of histone 3 of the exon 1_7 promoter (Weaver et al., 2004; Weaver et al., 2007). Thus, 5-HT activates both NGFI-A and the CREB-binding protein, which appear to associate with one another prior to DNA binding. The CREB-binding protein acetylates histones associated with the exon 1_7 promoter, enhancing the ability of NGFI-A to bind and activate gene transcription.

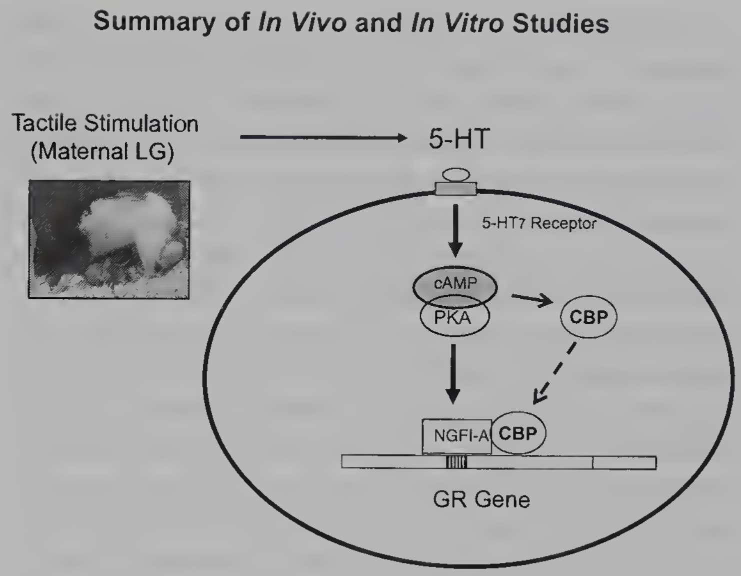

Summary of *In Vivo* and *In Vitro* Studies

Figure 3. A summary of in vivo studies with hippocampal tissue samples from neonates and in vitro studies using primary hippocampal cell cultures.
Note. In vivo, an increased frequency of pup LG associates with hippocampal 5-HT turnover, activation of a 5-HT7 receptor positively coupled to cAMP and cyclic nucleotide-dependent kinases (PKA), and the induction of NGFI-A expression. In vivo, increased pup LG or artificial tactile stimulation induces NGFI-A expression as well as that of the CREB-binding protein, both of which show greater binding to the exon 17 promoter in the neonatal offspring High- compared with low-LG mothers. Results of in vitro studies show that blockade of cAMP, PKA, or NGFI-A abolish the effect of 5-HT of glucocorticoid receptor expression. cAMP = concentration of adenosine 3,5-monophosphate; CREB-binding protein = cAMP-response element-binding protein; 5-HT = 5-hydroxytryptamine; LG = licking/grooming; NGFI-A = nerve growth factor-inducible factor A; PKA = protein kinase A.

Serotonin is a classic neurotransmitter that responds dynamically to environmental signals. This effect reflects the dependence of gene transcription on signals that derive from environmental events (note the relevant environmental event may be internal or external to the organism; e.g., a change in the availability of glucose, an electrical impulse, or a social interaction). Thus, the consequences of a variation in DNA sequence for phenotype will depend upon environmental signals. The genomic variant is relevant for phenotype only to the extent that the gene is actually expressed and participating in cellular function. If a particular tissue were to reveal a genomic variant in a region of the DNA that is transcriptionally inactive, then the implications for phenotype are minimal. The degree to which a genomic variant influences phenotype will depend upon the cellular context. Likewise, to the extent that an inherited genomic variant influences the capacity for transcription, then the influence of the environmental signal on phenotype will be affected by variation in nucleotide sequence. Such considerations showcase the interdependence of gene and environment.

Gene × Environment Interactions: The Case of the 5-HT Transporter

One of the best-studied genomic variations in biological psychiatry is that of the polymorphism in the promoter region of the *SLC6A4* gene that encodes for the 5-HT transporter (5-HTTP). The synaptic activity of 5-HT is terminated by reuptake into presynaptic terminals, which occurs via the 5-HTTP protein. Serotonin is a critical neurotransmitter in the regulation of emotional states (Lucki, 1998) and a major target for medications designed to treat a range of affective disorders (Blier & de Montigny, 1999; Gross & Hen, 2004). The gene for the 5-HTTP in humans shows a relatively common polymorphism characterized by a variable repeat sequence in the promoter region resulting in two common alleles: the short (*S*) variant comprising 14 copies of a 20–23 base-pair repeat unit, and the long (*L*) variant comprising 16 copies (Lesch et al., 1996).

The 5-HTTP promoter sequence polymorphism associates with differential transcription of the 5-HTTP gene, with more efficient expression from the *L* allele (Greenberg et al., 1999; Lesch et al., 1996). More recent studies suggest that further variation lying within the *L* allele also affects 5-HTTP gene function, resulting in subtypes of the *L* allele (Hu et al., 2005). The early studies of the functional correlates of the 5-HTTP promoter polymorphism as well as subsequent meta-analyses (Sen, Burmeister, & Ghosh, 2004; Schinka, Busch, & Robichaux-Keene, 2004; but also see Munafo, Clark, & Flint, 2005) reveal a statistical association with neuroticism, which is elevated among individuals that carry at least one copy of the *S* allele. Since neuroticism predicts an increased risk for affective disorders (Hettema, Neale, Myers, Prescott, & Kendler, 2006; Kendler, Kuhn, & Prescott, 2004), it is not surprising that the 5-HTTP promoter polymorphism is at least weakly associated with disorders of emotional function, namely, depression and anxiety (Canli & Lesch, 2007; Hariri, Emily, Drabant, & Weinberger, 2006).

Affective disorders associate with altered activity in the amygdala (Davis, 2006; Gorman, Kent, Sullivan, & Coplan, 2000; LeDoux, 2000; Meaney, LeDoux, & Leibowitz, 2008; Pérez-Edgar & Fox, 2005), a primary brain structure for emotional states of fear as well as a major target for 5-HT projections from the raphé nuclei. The expression of fear in rodents is inhibited by 5-HT activity in the hippocampus and amygdala (Stutzmann & LeDoux, 1999), and animals bearing mutations of 5-HT1A receptors show enhanced fearfulness (Gross & Hen, 2004). Aversive stimuli activate the amygdala in humans (LaBar et al., 1995; LaBar et al., 1998; Morris, Ohman, & Dolan, 1998) and individual differences in trait anxiety predict the response of the amygdala to threatening stimuli (Etkin, Egner, Peraza, Kande, & Hirsch, 2005; Etkin et al., 2004). Individuals carrying an *S* variant of the 5-HTTP promoter polymorphism show increased activation of the amygdala while processing fearful or angry faces (Hariri et al., 2002; Hariri et al., 2005; Pezawas et al., 2005). Similar effects of the 5-HTTP promoter polymorphism are obtained in subjects during processing of negative relative to neutral pictures (Heinz et al., 2005) or words (Canli et al., 2006). Moreover, the *S* allele of the 5-HTTP promoter is associated with increased activity at rest in both the amygdala and the hippocampus (Canli & Lesch, 2007).

An additional feature of anxiety is that of decreased inhibitory prefrontal regulation of evoked activity in the amygdala (Shin et al., 2004, 2005). Affective disorders associate with decreased volume in the subgenual region of the anterior cingulate cortex (Drevets et al., 1997), which projects to and regulates activity in the amygdala (Maren & Quirk, 2004). Gray matter volume in the subgenual anterior cingulate cortex is reduced in carriers of the *S* variant of the 5-HTTP promoter polymorphism. In addition, individuals bearing the *S* promoter variant show decreased functional connectivity between the amygdala and anterior cingulate cortex (Pezawas et al., 2005), which could serve as the basis for the increased amygdala response to negative stimuli.

Since genomic variations are potentially operative at any time over the lifespan, it is interesting to consider the possibility that the 5-HTTP polymorphism could alter 5-HT activity during periods of neural development. Thus, Pezawas et al. (2005) proposed that the association between the 5-HTTP promoter polymorphism and vulnerability to depression reflects a developmental process that affects the structural connectivity and functional interactions within a neural circuit that regulates emotional reactivity and fear. They further proposed that these functional deficits could be exacerbated by environmental adverse experiences, possibly through an impaired capacity to regulate affective states during periods of adversity. Such an effect would be expected to result in an interaction between gene and environment that renders individuals vulnerable to affective disorders. There is considerable evidence in support of this developmental hypothesis. Studies with rodent models reveal a profound influence of 5-HT activity on neuronal development, including the sprouting of axonal processes. Hen and colleagues (Gross et al., 2002) showed that a null mutation of the 5-HT1A in mice associates with increased fearfulness. This knockout model, like that of human sequence polymorphisms, reveals effects from conception onward to the time of behavioral assessment. It is therefore unclear as to whether the associated change in phenotype reflects a developmental influence, an effect on neurotransmission at the time of assessment, or both. Hen's lab (Gross & Hen, 2004; Gross et al., 2002) resolved this issue by producing a conditional knockout model in which the gene that encodes for the 5-HT1A receptor is rendered inoperative only during selected periods over the lifespan of the mouse. The findings reveal that the condition in which the 5-HT1A receptor gene is muted for only during the first 3 weeks of life completely recapitulates the increased fearfulness observed in animals in which the 5-HT1A receptor was knocked out for the entire life of the animal. These findings suggest that genomic variants that alter 5-HT function can act during early development to regulate neural circuitry that mediates the expression of fear. Moreover, variants of the gene for the 5-HTTP are associated with negative temperament in infancy (Auerbach et al., 2002; Ebstein, 2003, 2006), further suggesting that the consequences of 5-HT alterations appear in early life. Here again, however, there is environmental modulation of the genotype–phenotype relation. Fox et al. (2005) reported that the quality of the maternal environment interacted with the 5-HTTP promoter genotype to determine behavioral inhibition in children.

Fox et al.'s (2005) report is representative of a growing body of studies revealing Gene × Environment interactions in emotional and cognitive development. Suomi, Bennett, and colleagues described a brilliant example of such gene–environment interdependence in the rhesus monkey (Bennett et al., 2002; Champoux et al., 2002; Suomi, 2006). The rhesus monkey shows a polymorphism in the 5-HTTP promoter that is comparable in form and function to that in humans such that the *S* 5-HTTP promoter variant in the monkey is also associated with decreased 5-HTTP levels in brain (Lesch et al., 1997). Likewise, the 5-HTTP promoter polymorphism affects both 5-HT metabolism and emotional function in the monkey. Monkeys bearing the *S* version of the 5-HTTP promoter polymorphism show reduced 5-HT activity, increased impulsivity, and are more aggressive than animals bearing the longer version of the promoter variant (Bennett et al., 2002; Champoux et al., 2002). However, the genotype–phenotype relation is strongly influenced by the rearing environment. The phenotypic differences described earlier are apparent among animals separated from their mothers and reared in peer–peer groups. In contrast, among animals reared by their mothers, the differences in phenotype are actually slightly reversed such that animals with the *S* variant of the 5-HTTP promoter show modestly greater 5-HT activity and are less impulsive and aggressive than those carrying the longer promoter variant. In these studies, the genotype–phenotype relations depended completely upon the rearing environment. Similar effects are apparent in the mouse where there is an interaction between a disruption of the 5-HTTP gene and the quality of maternal care on neural systems associated with fear behavior (Carola et al., 2008).

A series of landmark studies (e.g., Caspi & Moffit, 2006; Caspi et al., 2003) provide a parallel set of findings in human populations. In these studies, as shown in previous research, childhood maltreatment increases the probability of an episode of major depression. However, this effect is only apparent among individuals that bear at least one copy of the *S* 5-HTTP promoter allele. Individuals homozygous for the *L* promoter variant show no increase in the probability of depression despite the experience of childhood maltreatment (see Caspi & Moffit, 2006, for a review of other, compelling examples of such Gene × Environment interactions). A more recent study shows evidence for the same Gene × Environment interaction in defining individual differences in neural functions that associate with depression (Canli et al., 2006). There is also evidence of a comparable Gene × Environment effect examining polymorphisms in other genes implicated in 5-HT neurotransmission (Jokela et al., 2007). In this case, the polymorphism in question is a SNP in the *5HTR2A* gene that encodes for the 5-HT2A receptor that is also linked to mood disorders. The authors found a significant interaction between the *5HTR2A* genotype and the quality of maternal care in defining the level of depressive symptoms.

The Caspi et al. (2003) finding of an interaction between the promoter polymorphism and either life stressors or childhood maltreatment has been replicated (Kaufman et al., 2004; Kendler, Kuhn, Vittum, Prescott, & Riley, 2005), and other studies suggest further moderation by gender (Eley et al., 2004; Grabe et al., 2005) or social support (Kaufman et al., 2004). Two studies (Gillespie, Whitfield, Williams, Heath, & Martin, 2005; Surtees et al., 2006) failed to replicate this Gene × Environment interaction effect. However, we need to appreciate that an interaction between a *single* genomic variant and a broadly characterized environmental condition at one stage of life is still a rather narrow focus on the complex pathophysiology of mental disorders. Genes encode for RNAs and proteins that operate in complex networks within cell signaling pathways. Such networks form the basis for the interaction of gene products that are then revealed as gene–gene interactions in studies of genomic variations, including data sets from genome-wide association studies. A major advance in genetics is the emergence of computational approaches that apply a biologically informed, statistical analysis of the interactions between functionally related genes. Such bioinformatic approaches feature the description of gene networks, involving multiple genomic targets and their variations. We can thus anticipate studies of Gene Network × Environment interactions.

Arguments for the interdependence of gene and environment should be familiar to those in developmental psychology. Belsky (1997) and Rutter (2007) argue from the other point of view that environmental forces, such as familial influences, will not affect all children equally and that an obvious source of differential sensitivity lies in genomic variation. In the nonhuman primate studies cited earlier, animals bearing the *L* variant of the 5-HTTP promoter polymorphism are markedly less sensitive to peer versus maternal rearing than those carrying the *S* promoter variant (at least with respect to the measures used in these studies). The same argument could be made for the Caspi et al. (2003) findings that reveal an increased risk for depression associated with child maltreatment among those bearing *S/S* 5-HTTP promoter variants. Indeed, one might well imagine that certain genomic variants alter neuronal function of the child, affecting the response to environmental conditions. Studies examining the children with a polymorphism on the *DRD4* gene that encodes a receptor for the neurotransmitter dopamine bear on this point. This genomic variant is associated with differences in attachment (Gervai et al., 2007; Lakatos et al., 2002) as well as infant responses to novelty (Lakatos et al., 2003). The *DRD4* polymorphism interacts with the quality of parenting to determine the level of internalizing and externalizing behavior in children (Propper, Willoughby, Halpern, Carbone, & Cox, 2007). Moreover, the same polymorphism determines the effect of a parent-training intervention targeting maternal sensitivity on the level of externalizing behavior in children (Bakermans-Kranenburg, Van IJzendoorn, Pijlman, Mesman, & Juffer, 2008). In these studies, the genotype of the child determined the response to parenting. Likewise, considering the evidence for the effects of genomic variants on emotional functions in children, genomic variation might operate within specific familial settings (e.g., Fox et al., 2005) to influence the behavior of the child and thus the quality of the interactions with caregivers. The 5-HTTP polymorphism associates with differences in infant temperament and reactions to novelty (Auerbach et al., 1999; Ebstein, 2006; Lakatos et al., 2003), which in turn predicts sensitivity to parental influences (Belsky, 1997). Although the point is less emphasized in this article, the Gene × Environment perspective argues no less forcefully that environmental effects must be considered within the genomic context of the individual. The interdependence of gene and environment is bidirectional.

These findings underscore the importance of Gene × Environment interactions, and the studies with the rhesus monkey suggest that the developmental consequences of such interactions persist into adulthood. This consideration simply reflects the long-standing interest of psychologists on the potential influence of early experience (Gottlieb, 1991, 1997), particularly those involving the interaction of the parent and child, on phenotypic development. In the preceding sections, we considered the molecular mechanisms that underlie Gene × Environment interactions with respect to immediate and transient changes in gene expression. But what are the mechanisms that might account for the enduring effects of environmental conditions on gene expression? How might parental influences stably alter phenotype? Could such effects involve the "programming" of gene expression? How might such stable influences occur? A response to such questions requires an understanding of the physical properties of Gene × Environment interactions that produce enduring effects on brain function.

Environmental Programming of Gene Expression

Studies in developmental psychobiology and physiology are replete with examples of the environmental programming of gene expression. Such studies commonly report that a variation in the early environment associates with changes in gene expression and function that persist into adulthood and thus well beyond the duration of the relevant environmental event. In the rat, for example, prenatal nutrient deprivation or enhanced exposure to hormonal signals associated with stress can stably alter, or program, the transcription of genes in the liver and other sites that are associated with glucose and fat metabolism, including the gene for the glucocorticoid receptor (Bateson et al., 2004; Gluckman & Hanson, 2004, 2007; Meaney, Szyf, & Seckl, 2007; Seckl & Holmes, 2007). These findings are assumed to represent instances in which the operation of a genomic region in adulthood varies as a function of early environmental influences. The results of recent studies suggest that such "programming" effects can derive from Gene × Environment interactions in early life that lead to a structural alteration of the DNA, which in turn mediates the effects on gene expression as well as more complex levels of phenotype (Meaney & Szyf, 2005; Meaney et al., 2007). These studies have been performed in rodents but were inspired by the vast literature reporting the pervasive effects of family environment on health outcomes in humans.

Epidemiological studies reveal the importance of family function and early life events as predictors of health in adulthood (Felitti et al., 1998; Leserman et al., 1996; Lissau & Sorensen, 1994; McCauley et al., 1997; Repetti, Taylor, & Seeman, 2002; Russak & Schwartz, 1997; Seckl, 2001). As adults, victims of childhood physical or sexual abuse, emotional neglect and harsh, inconsistent discipline are at considerably greater risk for mental illness, as well as for obesity, gastrointestinal illness, diabetes, and heart disease. "Stress diathesis" models are proposed as explanations for the relation between early experience and health (e.g., Gorman et al., 2000; Heim & Nemeroff, 2001; McEwen, 2003; Meaney, 2001; Meaney et al., 2007; Repetti et al., 2002; Seckl & Meaney, 1994). These models suggest that adversity in early life alters the development of neural and endocrine responses to stress in a manner that predisposes individuals to disease. Adversity or decreased quality of parental investment appears to directly increase the magnitude of emotional, autonomic, and endocrine responses to stress (collectively referred to here as *defensive responses*). Diathesis describes the interaction between development, including the potential influence of genetic variations at the level of sequence, and the prevailing level of stress in predicting health outcomes. Such models have considerable appeal and could identify both the origins of illness and the nature of the underlying vulnerability (as well as resilience). Moreover, such models have the virtue of providing a logical bridge between psychosocial (e.g., family function) and biological (endocrine responses to stressors) levels of analyses in understanding health outcomes.

A critical assumption in the stress diathesis models is that the increased expression of defensive responses endangers health. The idea has considerable support in clinical physiology (Chrousos & Gold, 1992; Dallman et al., 2005; McEwen, 2007). Defensive responses to stress are adaptive. Neural signals associated with the perception of the stressor increase the release of stress hormones into the bloodstream, including glucocorticoids from the adrenal gland and catecholamines, particularly norepinephrine, from the sympathetic nervous system. The combined actions of these hormones increase the availability of energy substrates, such as those derived from fat and glucose metabolism. Such effects maintain the normal cellular function and organ efficiency. These actions protect against catastrophes such as hypotensive shock. The release of catecholamines in the brain increases vigilance and provokes states of fear, and enhances avoidance learning and fear conditioning, which can serve to reduce the chances of further encounters with the same conditions.

Defensive responses are the logical outcome of the stress-induced changes in activity in the brain and endocrine organs, and are defined by increases in the synthesis and release of glucocorticoids and catecholamines (Dallman et al., 1995; Munck et al., 1984; Rosen & Schulkin, 1998; Sapolsky et al., 2000). However, there is a cost associated with persistent activation of these same responses: chronically enhanced emotional arousal, sustained increases in blood sugars and fats, possible hyperinflammation, disruption of sleep and normal cognitive function, among others (Chrousos & Gold, 1992; Dallman et al., 2005; McEwen, 2007; Sapolsky et al., 2000; Walker, 2007). Thus, chronic activation of defensive responses can indeed predispose individuals to illnesses such as diabetes, heart disease, mood disorders, and so on. Not surprisingly, studies in behavioral medicine reveal that individuals with enhanced stress reactivity are at greater risk for such forms of chronic illness. However, it is also important to note that insufficient activation of defensive responses under conditions of threat also compromises health and is associated with chronic fatigue, chronic pain, posttraumatic stress disorder, and hyperinflammation (Munck et al., 1984; Raison & Miller, 2003; Yehuda & Bierer, 2008). We walk a fine line here. And this underscores the importance of an appropriate level of stress reactivity for the individual, one sufficient to ensure the maintenance of function during adversity, but not so excessive as to promote chronic illness. What is appropriate, of course, will vary depending upon the prevailing level of environmental demand. There is no single, ideal level of stress reactivity across all populations.

Individual Differences in Defensive Responses

In the late 1950s and early 1960s, psychologists Gig Levine and Victor Denenberg (Denenberg, 1964; Levine, 1970) reported that postnatal handling (or infantile stimulation) of infant rodents decreased the magnitude of both behavioral and hypothalamic–pituitary–adrenal (HPA) responses to stress in adulthood. These findings demonstrated the influence of the early environment over the development of rudimentary defensive responses to threat. The importance of such findings in this era should not be underestimated. This was a period when defensive responses were considered as ''innate,'' developing well outside the realm of experiential influence. Levine and others later suggested that the effects of handling were actually mediated by changes in maternal care. Thus, handling of the pups was thought to alter the behavior of the mother, which was then critical for the ''handling'' effects. The handling paradigm involves exposure of the neonate to a complex set of stimuli, including that of a novel physical environment (Tang, Akers, Reeb, Romeo, & McEwen, 2006). Nevertheless, postnatal handling of rat pups does indeed increase the licking/grooming (LG) of pups by the mother (e.g., Lee & Williams, 1977; Liu et al., 1997). Pup LG is a major source of tactile stimulation for the neonatal rat that regulates endocrine and cardiovascular function in the pup (Hofer, 2005; Levine, 1994; Schanberg, Evoniuk, & Kuhn, 1984). The question then was whether such variations in pup LG might directly alter the development of individual differences in defensive responses.

Subsequent findings revealed considerable evidence for the effect of maternal care on the behavioral and endocrine responses to stress in the offspring. One approach was to simply examine the consequences of naturally occurring variations in pup LG among lactating rats independent of any experimental manipulation. Among lactating rats there are considerable individual differences in the frequency of pup LG that are stable over the reproductive lifetime of the female (Champagne, Francis, Mar, & Meaney, 2003). The results of longitudinal studies are consistent with the Levine (1970) hypothesis. The male or female adult offspring of mothers that naturally exhibit increased levels of pup LG (i.e., high-LG mothers) show more modest behavioral and endocrine responses to stress compared to animals reared by low-LG mothers (Caldji et al., 1998; Francis, Diorio, Liu, & Meaney, 1999; Liu et al., 1997; Menard, Champagne, & Meaney, 2004; Toki et al., 2007; Weaver et al., 2004; Zhang et al., 2006). Specifically, the offspring of high-LG mothers show reduced fearfulness and more modest HPA responses to stress. Cross-fostering studies, where pups born to high-LG mothers are fostered at birth to low-LG mothers (and vice versa), suggest a direct relation between maternal care and the postnatal development of individual differences in behavioral and HPA responses to stress (Caldji, Diorio, & Meaney, 2003; Caldji, Francis, Sharma, Plotsky, & Meaney, 2000; Francis et al., 1999; Weaver et al., 2004). In these studies, the rearing mother determined the phenotype of the offspring. Thus, variations within a normal range of parental care can dramatically alter phenotypic development in the rat.

The effects of maternal care on the development of defensive responses to stress in the rat involve alterations in the function of the corticotrophin-releasing factor (CRF) systems in selected brain regions (Figure 4). The CRF system furnishes the critical signal for the activation of behavioral, emotional, autonomic, and endocrine responses to stressors (Bale & Vale, 2004; Koob, Heinrichs, Menzaghi, Pich, & Britton, 1994; Plotsky, Cunning-

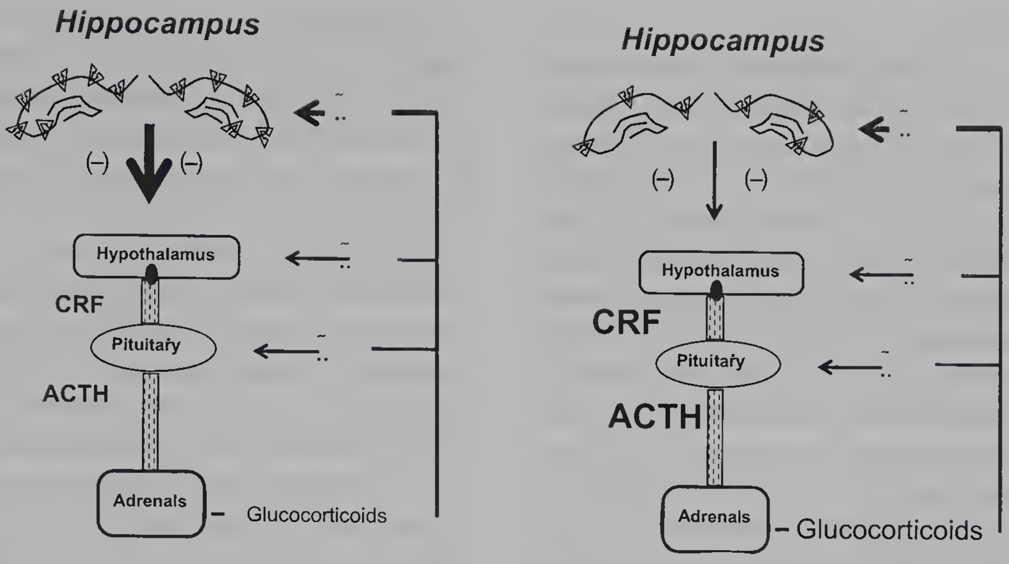


Figure 4. A schema outlining the function of the hypothalamic–pituitary–adrenal axis, the nexus of which are the corticotropin-releasing factor (CRF) neurons of the paraventricular nucleus of the hypothalamus.
Note. CRF is released into the portal system of the anterior pituitary stimulating the synthesis and release of adrenocorticotropin (ACTH), which then stimulates adrenal glucocorticoid release. Glucocorticoids act on glucocorticoid receptors in multiple brain regions, including the hippocampus, to inhibit the synthesis and release of CRF (i.e., glucocroticoid negative feedback). The adult offspring of high-LG mothers, by comparison to those of low-LG dams, show (a) increased glucocorticoid receptor expression, (b) enhanced negative-feedback sensitivity to glucocorticoids, (c) reduced CRF expression in the hypothalamus, and (d) more modest pituitary–adrenal responses to stress. LG = licking/grooming.

ham, & Widmaier, 1989). As adults, the offspring of high-LG mothers show decreased CRF expression in the hypothalamus, as well as reduced plasma adrenocorticotropin (ACTH) and glucocorticoid responses to acute stress by comparison to the adult offspring of low-LG mothers (Francis et al., 1999; Liu et al., 1997; Weaver et al., 2004; Weaver et al., 2005). Circulating glucocorticoids act at glucocorticoid receptor sites in corticolimbic structures, such as the hippocampus, to regulate HPA activity (Figure 4). Such feedback effects commonly inhibit hypothalamic CRF expression. The high-LG offspring showed significantly increased hippocampal glucocorticoid receptor expression, enhanced glucocorticoid negative-feedback sensitivity, and decreased hypothalamic CRF levels. Indeed, the magnitude of the glucocorticoid response to acute stress was significantly correlated with the frequency of pup LG during the 1st week of life, as was the level of both hippocampal glucocorticoid receptor and hypothalamic CRF expression (all *r*s > .70; Liu et al., 1997). Importantly, pharmacological manipulations that block the effect of the glucocorticoid receptor eliminate the maternal effect on the HPA response to stress, suggesting that the differences in hippocampal glucocorticoid receptor expression are directly related to those at the level of HPA function.

Pup LG is a major source of tactile stimulation for the neonate. Experimental models that directly apply tactile stimulation, through the stroking of the pup with a brush, provide direct evidence for the importance of tactile stimulation derived from pup LG. Thus, stroking pups over the 1st week of life increases hippocampal glucocorticoid receptor expression (Jutapakdeegul, Casalotti, Govitrapong, & Kotchabhakdi, 2003) and dampens behavioral and HPA responses to stress (Burton et al., 2007; Gonzalez, Lovic, Ward, Wainwright, & Fleming, 2001). Likewise, manipulations of lactating mothers that directly increase the frequency of pup LG also increase hippocampal glucocorticoid receptor expression and decrease HPA responses to stress (Francis et al., 1999; Toki et al., 2007).

The offspring of the high- and low-LG mothers also differed in behavioral responses to stress (Caldji et al., 1998; Caldji et al., 2003; Francis et al., 1999). As adults, the offspring of the high-LG mothers showed decreased startle responses, substantially less fearfulness in the presence of stressors, such as novel environments. Moreover, active defensive responses, such the burying of threatening stimuli, are more conspicuous among the adult offspring of low-LG mothers (Menard et al., 2004). While such pervasive effects are no doubt associated with alterations in multiple neural systems, there are robust effects of maternal care on the central CRF systems, including those that lie outside of the hypothalamus. The activation of fear behavior corresponds to an increase in CRF release from the amygdala and bed nucleus of the stria terminalis onto catecholaminergic cell bodies in the locus coeruleus. CRF acts at CRF$_1$ receptors in the locus coeruleus to stimulate the release of norepinephrine in a variety of corticolimbic structures (Bale & Vale, 2004; Valentino, Curtis, Page, Pavcovich, & Florin-Lechner, 1998). Gamma-aminobutyric acid type A (GABA), which is the primary source of neural inhibition in the adult mammalian brain, dampens the activation of this CRF–catecholamine connection during stress. The offspring of the high-LG mothers show decreased CRF$_1$ receptor levels in the locus coeruleus and increased GABA$_A$/benzodiazepine receptor levels in the basolateral and central nucleus of the amygdala, as well as in the locus coeruleus and decreased CRF expression in the amygdala (Caldji et al., 1998; Caldji et al., 2003). The adult offspring of high-LG mothers are more sensitive to the inhibitory effects of benzodiazepines on fear behavior (Fries, Moragues, Caldji, Hellhammer, & Meaney, 2004). Receptors for the benzodiazepines (e.g., diazepam) are part of the GABA$_A$ receptor complex, and the anxiolytic effects of these compounds occur through an ability to enhance the effect of GABA at the GABA$_A$ receptor site. GABA$_A$, or benzodiazepine receptor agonists, suppress CRF expression in the amygdala, thus reducing activity within the amygdala–locus coeruleus and decreasing norepinephrine responses to stress. Predictably, stress-induced increases in norepinephrine that are normally stimulated by CRF are significantly higher in the offspring of the low-LG offspring. The increased release of norepinephrine is consistent with the enhanced fearfulness of these animals. Thus, increased pup LG is associated with the enhanced efficacy of systems that normally serve to inhibit the expression and actions of CRF. These systems include the hippocampal glucocorticoid receptor and the GABA$_A$ receptor in the amygdala. Both effects involve sustained alterations in gene expression as a function of maternal care.

The complexity of such maternal effects on gene expression is apparent in the alterations in GABA$_A$ receptor function. The GABA$_A$ receptor is a multiprotein complex that comprised five individual subunits, each of which is a product of a distinct genomic region. An interesting feature to this system is that there are at least 19 different subunits that can be employed in the formation of a GABA$_A$ receptor. The activity of the receptor is determined by its subunit composition. For example, the inclusion of an a$_1$ subunit confers an increased affinity of the receptor for GABA, enhancing GABAergic activity at the GABA$_A$ receptor. Others, such as the g$_2$ subunit, will define the presence of a benzodiazepine binding site that further increases the inhibitory function of the GABA$_A$ receptor. The adult offspring of high-LG mothers show significantly increased expression of both the a$_1$ and g$_2$ subunits. The effect is almost unique to the amygdala and is reversed with cross-fostering (Caldji et al., 2003).

The results of these studies suggest that the behavior of the mother toward her offspring can "program" stable changes in gene expression that then serve as the basis for individual differences in behavioral and neuroendocrine responses to stress in adulthood. The maternal effects on phenotype are associated with sustained changes in the expression of genes in brain regions that mediate responses to stress, and form the basis for stable individual differences in stress reactivity. These findings provide a potential mechanism for the influence of parental care on vulnerability/resistance to stress-induced illness over the lifespan. However appealing, this hypothesis has yet to be directly confirmed. But the critical issue for this article is simply that of *how* maternal care might stably affect gene expression. How are variations in the social interactions between the mother and her offspring "biologically embedded" so as to stably alter the activity of specific regions of the genome? The answers to these questions appear to involve the ability of social interactions in early development to actually modify the structure of the relevant genomic regions.

Epigenetic Regulation of the Genome

When we think of genomic influences we most commonly imagine effects associated with variation in nucleotide sequence—the genetic code. Yet, this is only one form of information contained on the DNA. Despite the reverence afforded DNA, a gene is basically like any other molecule in the cell; it is subject to physical modifications. These modifications alter the structure and chemical properties of the DNA, and thus gene expression. Collectively, the modifications to the DNA and its chromatin environment can be considered as an additional layer of information that is contained within the genome. This information is thus *epigenetic* in nature (the name derives from the Greek *epi* meaning "upon" and *genetics*). The acetylation of the histone proteins referred to earlier is one example of an epigenetic modification. Epigenetic modifications do not alter the sequence composition of the genome. Instead, epigenetic marks on the DNA and the other features of the chromatin regulate the operation of the genome. Thus, *epigenetics* has been defined as a functional modification to the DNA that does not involve an alteration of sequence. While this definition has recently been subjected to revision (Bird, 2007; Hake & Allis, 2006), the essential features of epigenetic mechanisms are (a) structural modifications to chromatin either at the level of the histone proteins (Figure 2) or the DNA, (b) regulation of the structure and function of chromatin, (c) affects on gene expression, and (d) that these effects occur in the absence of any change in nucleotide sequence. The functional byproduct of the epigenetic modifications is that of a change in gene transcription.

The classic epigenetic alteration is that of DNA methylation, which involves the addition of a methyl group onto cytosines in the DNA (Bird, 1986; Holliday, 1989; Razin & Cedar, 1993; Razin & Riggs, 1980). The methylation of DNA is an active biochemical modification that in mammals selectively targets cytosines and is achieved through the actions of a class of enzymes, DNA methyltransferases, which transfer the methyl groups from methyl donors. There are two critical features to DNA methylation: First, it is a stable chemical modification, and second, it is associated with the silencing of gene transcription (Bestor, 1998; Bird, 2002; Bird & Wolffe, 1999; Razin, 1998).

Until recently, DNA methylation patterns on the genome were thought to be overlaid upon the genome only during early periods in embryonic development. This belief was derived in part from the experimental models commonly used to study DNA methylation. DNA methylation-induced gene silencing mediates two of the most commonly studied examples of epigenetic silencing, namely, X-chromosome inactivation and gene imprinting. Mammalian females bear two copies of the X-chromosome. The inactivation of one copy of the X-chromosome occurs in all mammalian females and is essential for normal function (i.e., maintaining a constant gene dosage in males and females). The silencing of the X-chromosome is associated with DNA methylation (Mohandas, Sparkes, & Shapiro, 1981; Riggs & Pfeifer, 1992; see Hellman & Chess, 2007, for a more current update). The second example of epigenetic-mediated gene silencing is that of gene imprinting (da Rocha & Ferguson-Smith, 2004; Reik, 2001), a remarkable subject in its own right, and one with considerable implications for growth and development (Charalambous, da Rocha, & Ferguson-Smith, 2007). For humans and other mammals, the expression-specific genes are determined by the parent of origin. For certain genes, the copy derived from the mother is active, while that emanating from the father is silenced—a ''maternally imprinted gene.'' In other cases, it is the reverse; the copy of the gene inherited from the father that is active, while that from the mother is silenced—a ''paternally imprinted gene.'' The silent copy is methylated in DNA regions that regulate gene expression and thus inactive. Again, the epigenetic marks associated with gene imprinting are established very early in life. These marks, as well as those associated with X-chromosome inactivation, are stable, leaving researchers in the field with the impression that under normal conditions DNA methylation occurs early in embryonic life and is largely irreversible. Indeed, it was commonly thought that DNA methylation was an actively dynamic process only during periods of cell division such that in mature, postmitotic cells further alteration of methylation patterns was improbable. Moreover, the loss of cytosine methylation in such models is associated with profound pathology. This perspective was further reinforced by findings showing that an alteration of DNA methylation at critical genomic targets (i.e., tumor suppressors) is associated with cancer (Eden, Gaudet, Waghmare, & Jaenisch, 2003; Feinberg, 2007; Laird, 2005).

While these assumptions concerning DNA methylation appear valid for the examples cited earlier, recent studies reveal that DNA methylation patterns are actively modified in mature (i.e., fully differentiated) cells, including, and perhaps especially, neurons and that such modifications can occur in animals in response to cellular signals driven by environmental events (Bird, 2007; Jirtle & Skinner, 2007; Meaney & Szyf, 2005). For example, variations in the diet of mice during gestation or later in development, such as the early postweaning period, can stably alter the methylation status of the DNA (Cooney, Dave, & Wolff, 2002; Waterland & Jirtle, 2003; Waterland, Lin, Smith, & Jirtle, 2006; Whitelaw & Whitelaw, 2006). Likewise, both mature lymphocytes (Bruniquel & Schwartz, 2003; Murayama et al., 2006) and neurons (e.g., Champagne, 2008; Champagne et al., 2006; Lubin, Roth, & Sweatt, 2008; Martinowich et al., 2003; Sweatt, 2009) show changes in the DNA methylation patterns at critical genomic regions in response to environmental stimuli that stably alter cellular function. The ability of environmental signals to actively remodel epigenetic marks that regulate gene expression is a rather radical change in our understanding of the environmental regulation of gene expression. Such epigenetic modifications are thus a candidate mechanism for the environmental ''programming'' of gene expression.

DNA Methylation and Gene Transcription

DNA methylation is associated with the silencing of gene transcription. This effect appears to be mediated in one of two ways (Bird, 2002). First, wide swaths of DNA can be methylated and the shear density of methylation precludes transcription factor binding to DNA sites, thus silencing gene expression. The second manner is subtler, and probably far more prevalent, in regions with more dynamic variations in gene transcription, such as the brain. In this case, selected cytosines are methylated and the presence of the methyl group attracts a class of proteins know as methylated-DNA binding proteins (Klose & Bird, 2006). These proteins, in turn, attract an entire cluster of proteins, known as the repressor complexes that are the active mediators of the gene silencing. The HDACs are a critical component of the repressor complex. HDACs prevent histone acetylation and favor a closed chromatin state that constrains transcription factor binding and gene expression (Figure 2 and see earlier). Compounds that inhibit HDACs can thus increase transcription from methylated DNA.

Epigenetics and the Social Environment

The following section describes studies of the molecular basis for the effects of maternal care on the development of individual differences in gene expression and stress responses. The mechanism for this interaction is epigenetic, involving alterations in DNA methylation at specific sites in the genome. In summary, variations in mother–infant interactions in the rat alter the extra- and intracellular environment of neurons in selected brain regions. Such alterations directly modify the epigenetic marks on regions of the DNA that regulate the transcription of the glucocorticoid receptor, which in turns regulates the HPA response to stress. These epigenetic marks are stable, enduring well beyond the period of maternal care, and thus provide a molecular basis for a stable maternal effect on the phenotype of the offspring. Thus, the behavior of the mother directly alters cellular signals that then actively sculpt the epigenetic landscape of the offspring, influencing the activity of specific regions of the genome and the phenotype of the offspring.

Epigenetic Effects of Variations in Maternal Care

The critical feature of the maternal effects described earlier is that of persistence. The differences in the frequency of pup LG between high- and low-LG mothers are limited to the 1st week of postnatal life. And yet the differences in gene expression and neural function are apparent in adulthood. How might the effects of an essentially social interaction stably alter the expression of the glucocorticoid receptor gene?

The focus of the epigenetic studies is the NGFI-A consensus sequence in the exon 1_7 promoter (Figure 1) that activates glucocorticoid receptor expression in hippocampal neurons. The tactile stimulation associated with pup LG increases 5-HT metabolism in the hippocampus. As described earlier (Figure 3), in vitro studies show that 5-HT acts on 5-HT$_7$ receptors to initiate a series of intracellular signals that culminate with an increase in the expression of NGFI-A as well as in the CREB-binding protein. Comparable effects occur in vivo. Manipulations that increase pup LG by lactating rats result in an increased level of cAMP as well as NGFI-A (Meaney et al., 2000). Pups reared by high-LG mothers show increased NGFI-A expression in hippocampal neurons as well as an increased binding of NGFI-A to the exon 1_7 promoter sequence (Hellstrom , Zhang, Diorio, & Meany, 2009; Weaver et al., 2007). Moreover, the binding of NGFI-A to the exon 1_7 promoter sequence is actively regulated by mother–pup interactions, such that there is increased NGFI-A bound to the exon 1_7 promoter immediately following a nursing bout but not at a period that follows 25 min without mother–pup contact (Hellstrom et al., 2009).

Nerve growth factor-inducible factor A and the CREB-binding protein form a complex that binds directly to the exon 1_7 promoter sequence and actively redesigns the methylation pattern at this region of the genome (Weaver et al., 2004; Weaver et al., 2007). Thus, as adults the offspring reared by high-LG mothers show very modest levels of methylation at the 5′-CpG of the NGFI-A consensus sequence (Figure 5). This effect on methylation is very precise. Lying only a few nucleotides removed from this site is 3′-CpG site (Figure 5), the methylation status of which is unaffected by maternal care.

A rather novel aspect to the effect of maternal care on DNA methylation was apparent in the results of a simple developmental study examining the methylation status of the 5′- and 3′-CpG sites from late in fetal life to adulthood (Weaver et al., 2004). Neither the 5′- nor the 3′-CpG site is methylated in hippocampal neurons from fetal rats, whereas both sites are heavily methylated on the day following birth, with no difference as a function of maternal care. These findings reflect what is

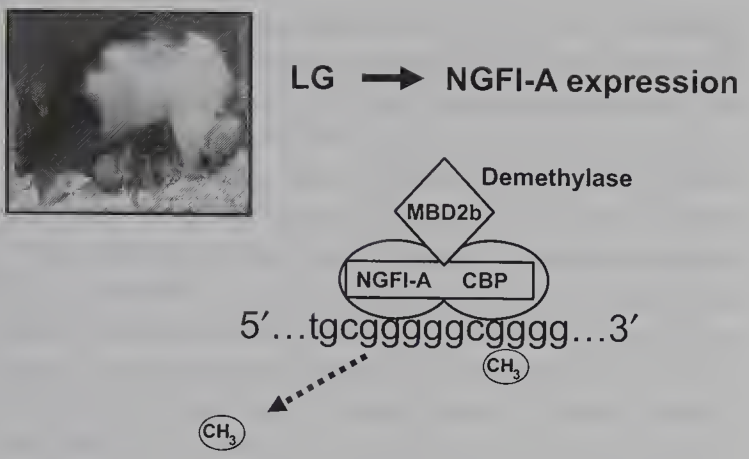


Figure 5. A working hypothesis for the experience (maternal care) driven remodeling of the epigenetic state of the NGFI-A consensus binding sequence over the 1st week of postnatal life in the offspring of high-LG mothers.
Note. The binding of a NGFI-A/CBP complex actively targets the association of a putative demethylase (MBD2) resulting in the removal of the methyl group from the 5′-CpG site of the NGFI-A binding site (Meaney & Szyf, 2005). CBP = CREB-binding protein; LG = licking/grooming; NGFI-A = nerve growth factor-inducible factor A.

referred to as *de novo methylation*, whereby a methyl group is applied to previously unmethylated sites. However, between the day following birth and the end of the 1st week of life, the 5′-CpG is "demethylated" in pups reared by high-, but not low-LG mothers. The difference then persists into adulthood. Importantly, the period over which the demethylation occurs is precisely that during which high- and low-LG mothers differ in the frequency of pup LG; the difference in pup LG between high- and low-LG mothers is not apparent in the 2nd week of postnatal life (Caldji et al., 1998; Champagne, Francis, et al., 2003).

The demethylation of the 5′-CpG site occurs as a function of the same 5-HT-activated signals that regulate glucocorticoid receptor gene expression in cultured hippocampal neurons (Weaver et al., 2007). Thus, when hippocampal neurons of embryonic origin are placed in culture and treated with 5-HT, which mimics the extracellular signal associated with maternal LG, the 5′-CpG site is demethylated; there is no effect at the 3′-CpG site. The binding of NGFI-A to the exon 1_7 site is critical. Hippocampal neurons that are rendered incapable of increasing NGFI-A expression through antisense or siRNA treatment show neither the demethylation of the 5′-CpG site nor the increase in glucocorticoid receptor expression (Weaver et al., 2007). Likewise, a mutation of the NGFI-A site that completely abolishes the binding of NGFI-A to the exon 1_7 promoter also prevents the demethylation of the 5′-CpG. Finally, the infection of hippocampal neurons with a virus containing a construct that was engineered to express high levels of NGFI-A produces demethylation of the 5′-CpG of the exon 1_7 promoter sequence and increased glucocorticoid receptor expression.

But there is a complication. If DNA methylation blocks transcription factor binding and the 5′-CpG site of the exon 1_7 promoter is heavily methylated in neonates, then how might maternally activated NGFI-A bind to and remodel the exon 1_7 region? And why is the effect apparent at the 5′- but not the 3′-CpG? The answer to these questions appears to involve other transcriptional signals that are affected by maternal care. Levels of the transcription factor specific protein-1 (SP-1) and the CREB-binding protein are also increased in the hippocampus of pups reared by high-LG mothers (Hellstrom et al., 2009; Weaver et al., 2007). The exon 1_7 promoter contains a DNA sequence that binds SP-1 and this region overlaps with that for NGFI-A. SP-1 can actively target both methylation and demethylation of CpG sites (Brandeis et al., 1994). The 5′-CpG site is the region of overlap in the binding sites. The CREB-binding protein, on the other hand, acts as a histone acetyltransferase, an enzyme capable of acetylating histone tails, including the exon 1_7 region, opening chromatin and permitting the binding of transcription factors such as NGFI-A and SP-1. Increasing histone acetylation can lead to transcription factor binding at previously methylated sites, and the subsequent demethylation of these regions (Szyf, Weaver, Champagne, Diorio, & Meaney, 2005).

The NGFI-A transcription factor binds to multiple sites across the genome. If NGFI-A-related complexes target demethylation, then one might assume that other NGFI-A-sensitive regions should show a maternal effect comparable to that observed with the glucocorticoid receptor. Zhang, Hellstrom, Wei, and Meaney (2009) showed that the hippocampal expression of the *GAD1* gene that encodes for glutamic acid decarboxylase, a rate limiting enzyme in the production of GABA, is increased in the adult offspring of high-LG mothers. This effect is associated with increased cytosine methylation of an NGFI-A response element. Moreover, as with the effect on the glucocorticoid receptor, an in vitro increase in NGFI-A expression mimics the effects of increased pup LG.

In summary, the maternally induced changes in specific intracellular signals in hippocampal neurons can physically remodel the genome. The increased binding of NGFI-A that derives from pup LG appears critical for the demethylation of the exon 1_7

promoter. We suggest that this process involves accompanying increases in SP-1 and the CREB-binding protein, and that the combination of these factors results in the active demethylation of the exon 1_7 promoter. It should be noted that there are important features of this model that remain to be clearly defined, including the identification of the enzyme that is directly responsible for the demethylation. Nevertheless, the events described to date represent a model by which the biological pathways activated by a social event may become imprinted onto the genome. This imprint is then physically apparent in the adult genome, resulting in stable alterations (or programming) of gene expression.

The Functional Importance of the Social Imprint

The presence of a methyl group on the 5′-CpG of the NGFI-A binding site is functionally related to glucocorticoid receptor gene expression in adult animals. In vitro studies reveal that the methylation of the 5′-CpG site reduces the ability of NGFI-A to bind to the exon 1_7 promoter and activate glucocorticoid receptor transcription (Weaver et al., 2007). These findings are consistent with the model described earlier, whereby DNA methylation impedes transcription factor binding and thus the activation of gene expression. The question concerns the in vivo situation. In contrast to the situation with neonates, there is no difference in NGFI-A expression as a function of maternal care among adult animals. However, the altered methylation of the exon 1_7 promoter would suggest differences in the access of NGFI-A to its binding site on the promoter. Chromatin-immunoprecipitation assays, which permit measurement of the direct interaction between a specific protein and a defined region of the DNA, reveal increased NGFI-A association with the exon 1_7 promoter in hippocampi from adult offspring of high- compared to low-LG mothers (Weaver et al., 2004; Weaver et al., 2005). These findings show that in the living animal, under normal conditions, there is more NGFI-A associated with the exon 1_7 promoter in hippocampal neurons of adult animals reared by high- compared with low-LG mothers.

There is also evidence that directly links the maternal effect on the epigenetic state of the exon 1_7 promoter to the changes in glucocorticoid receptor expression and thus HPA responses to stress. Recall that the methylation of specific CpG sites can diminish transcription factor binding through the recruitment of repressor complexes that include HDACs. The HDACs deactylate histone tails, thus favoring a closed chromatin configuration. Indeed, the exon 1_7 promoter is more prominently acetylated in hippocampi from adult offspring of high- compared with low-LG mothers (Weaver et al., 2004, 2005). This finding is consistent with the increased transcription of the glucocorticoid receptor gene in animals reared by high- versus low-LG mothers. A subsequent study (Weaver et al., 2004) examined the effects of directly blocking the actions of the HDACs in the adult offspring of high- and low-LG mothers. An HDAC inhibitor was infused directly into the hippocampus daily for 4 consecutive days. The treatment with the HDAC inhibitor produces a series of predictable results that reflect a cause–effect relation between DNA methylation and gene expression. First, HDAC blockade eliminates the differences in the acetylation of the histone tails of the exon 1_7 promoter in hippocampal samples from high- and low-LG mothers. Second, the increased histone acetylation of the exon 1_7 promoter in the offspring of low-LG mothers is associated with an increase in the binding of NGFI-A to the exon 1_7 promoter in the offspring of low-LG mothers, eliminating the maternal effect on NGFI-A binding to the exon 1_7 promoter. Comparable levels of NGFI-A binding to the exon 1_7 promoter then eliminate the maternal effect on hippocampal glucocorticoid receptor expression, such that glucocorticoid receptor levels in the adult offspring of low-LG mothers treated with the HDAC inhibitor are comparable to those in animals reared by high-LG mothers. And most importantly, the infusion of the HDAC inhibitor reversed the differences in the HPA response to stress.

Histone deacetylase inhibition increases NGFI-A binding to the exon 1_7 promoter in the offspring of low-LG mothers. The studies with neonates reveal that increased NGFI-A binding results in the demethylation of the 5′-CpG. In vitro, the introduction of a viral tool that leads to the increased expression of NGFI-A is sufficient to demethylate the exon 1_7 promoter. We (Weaver et al., 2007) argue that the binding of NGFI-A is critical for the demethylation of the 5′-CpG site. The same effect is apparent in vivo and even with the adult animals used in the studies described earlier. HDAC infusion into the hippocampus increases NGFI-A binding to the exon 1_7 promoter in the adult offspring of low-LG mothers and decreases the level of methylation of the 5′-CpG site on the exon 1_7 promoter. A subsequent study (Weaver et al., 2005) showed that the reverse pattern of results could be obtained in response to a dietary manipulation (methionine):

Greater methylation of the 5′-CpG in the offspring of high-LG mothers decreased NGFI-A binding and GR expression and increased HPA responses to stress (Weaver et al., 2005).

While these studies employ rather crude pharmacological manipulations, the results are critical as they suggest that fully mature neurons in an adult animal express the necessary enzymatic machinery to demethylate or remethylate DNA. Thus, it is possible that environmentally driven changes in neuronal transcriptional signals could potentially remodel the methylation state of specific regions of the DNA (Meaney & Szyf, 2005). The cytosine methylation state of a promoter for the brain-derived neurotrophic factor (*bdnf*) gene is also influenced by maternal care in early life (Roth, Lubin, Funk, & Sweatt, 2009). Indeed, Lubin et al. (2008) provided evidence for an alteration of the methylation state of the same *bdnf* promoter following contextual fear conditioning in adult rats (also see Bredy et al., 2007). These effects are consistent with previous reports of activity-dependent alterations in the methylation of the same *bdnf* promoter (Martinowich et al., 2003) and suggest that epigenetic states might be altered by a wide range of biologically relevant events that result in synaptic remodeling (Meaney & Szyf, 2005; Renthal & Nestler, 2008; Sweatt, 2009). Such epigenetic modifications might therefore underlie a wide range of stable changes in neural function following exposure to highly salient events (e.g., chronic stress, drugs of abuse, reproductive phases such as parenting, etc.), and are thus logical mechanisms for environmentally induced alterations in mental health (Akbarian & Huang, 2009; Jiang et al., 2008; Tsankova, Renthal, Kumar, & Nestler, 2007). While such effects have yet to be reported for DNA methylation, modifications of histone proteins are associated with exposure to drugs of abuse and stressors in rodent models (Renthal & Nestler, 2008; Renthal et al., 2007).

A set of recent studies (McGowan et al., 2009) suggests that comparable epigenetic modifications might occur in humans in response to variations in parent–offspring interactions. DNA was extracted from hippocampal samples obtained from victims of suicide or from individuals that had died suddenly from other causes (auto accidents, heart attacks, etc.). The samples were obtained from the Québec Suicide Brain Bank, which conducts forensic phenotyping that includes a validated assessment of psychiatric status and developmental history (e.g., McGirr, Renaud, Seguin, Alda, & Turecki, 2008). The studies examined the methylation status of the exon 1_F promoter of the glucocorticoid receptor, which corresponds to the exon 1_7 promoter in the rat (Turner & Muller, 2005). The results showed increased DNA methylation of the exon 1_F promoter in hippocampal samples from suicide victims compared with controls, but only if suicide was accompanied with a developmental history of child maltreatment. Child maltreatment, independent of psychiatric state, predicted the DNA methylation status of the exon 1_F promoter. As in the previous rodent studies, the methylation state of the exon 1_F promoter also determined the ability of NGFI-A to bind to the promoter and activate gene transcription. While such studies are obviously correlational, and limited by postmortem approaches, the results are nevertheless consistent with the hypothesis that variations in parental care can modify the epigenetic state of selected sites of the human genome. Moreover, the findings are also consistent with studies that link childhood abuse to individual differences in stress responses (Heim et al., 2000). Childhood abuse was associated with an increase in pituitary ACTH responses to stress among individuals with or without concurrent major depression. Heim et al.'s (2000) findings are particularly relevant since pituitary ACTH directly reflects central activation of the HPA stress response and hippocampal GR activation dampens HPA activity. These findings are consistent the rodent studies cited earlier investigating epigenetic regulation of the glucocorticoid receptor gene and with the hypothesis that early life events can alter the epigenetic state of relevant genomic regions, the expression of which may contribute to individual differences in the risk for psychopathology (Holsboer, 2000; Neigh & Nemeroff, 2006; Schatzberg, Rothschild, Langlais, Bird, & Cole, 1985).

Transgenerational Effects

Individual differences in stress responses in the adult rat are associated with naturally occurring variations in maternal care during infancy. Manipulations that alter mother–pup interactions in the rat alter patterns of gene expression and stress response in the offspring (e.g., Francis et al., 1999; Meaney, 2001; Roth et al., 2009; Toki et al., 2007). Such effects are certainly familiar to child psychologists working with parenting training programs. Moreover, these maternal effects might also serve as a possible nongenomic mechanism by which selected traits could be transmitted from one generation to another through variations in mother–pup interactions. Interestingly, low-LG mothers are more fearful than are high-LG dams (Francis, Champagne, & Meaney, 2000). Individual differences in stress reactivity are apparently transmitted across generations: Fearful mothers beget more stress reactive offspring. The obvious question is whether the transmission of these traits occurs only as a function of genomic-based inheritance. If this is the case, then the differences in maternal behavior may be simply be an epiphenomenon and not causally related to the development of individual differences in stress responses. The issue is not one of inheritance but rather the mode of inheritance. However, the results of the previously cited cross-fostering studies indicate that individual differences at the level of gene expression or complex phenotype can be directly altered during the postnatal period by maternal behavior. This interpretation is buttressed from studies showing that tactile stimulation of pups by the experimenter with a brush, so-called stroking, increases hippocampal NGFI-A (Hellstrom et al., 2009) increases hippocampal GR expression (Jutapakdeegul et al., 2003), and dampens HPA responses to stress in adulthood (Burton et al., 2007; Gonzalez et al., 2001).

The cross-fostering studies also reveal that individual differences in maternal behavior are transmitted from mother to the female offspring. Hence, as adults, the female offspring of more fearful, low-LG mothers are also more fearful low-LG mothers (Francis et al., 1999; also see Fleming, O'Day, & Kraemer, 1999). Regardless of their biological origins, females that are reared by high-LG mothers are less fearful and show an increased frequency of pup LG. And the mechanism for the intergenerational transmission of such individual differences appears to be the difference in the frequency of pup LG during early postnatal life. Fleming and colleagues found that as adults, females rats deprived of maternal showed reduced frequencies of multiple forms of maternal behavior, and such effects were reversed if animals were provided with tactile stimulation by stroking the pups with a brush over the first weeks of life (Gonzalez et al., 2001; Lovic, Gonzalez, & Fleming, 2001; Melo et al., 2006). Moreover, the stroking in infancy also increased the expression of maternal behaviors in response to ovarian hormones in adulthood (Novakov & Fleming, 2005), an effect that is comparable to that associated with increased pup LG (Champagne, Diorio, Sharma, & Meaney, 2001; see also next).

The individual differences in pup LG involve differences in estrogen receptor α-gene expression in the medial preoptic area (MPOA) of the hypothalamus, a region that is critical for maternal behavior in the rat (Fleming et al., 1999; Numan & Insel, 2003). There is increased estrogen receptor a expression in female offspring of high-LG mothers (Champagne, Weaver, Diorio, Sharma, & Meaney, 2003). Estrogen acts during late gestation to increase oxytocin receptor levels in the MPOA (Fahrbach & Pfaff, 1986; Pedersen, 1997), and this effect is greater in the female offspring of high- compared to low-LG mothers (Champagne et al., 2001), reflecting the increased sensitivity to estrogen. Oxytocin appears to act at oxytocin receptor levels in the MPOA to facilitate the release of dopamine from neurons in the ventral tegmental nucleus, and the increased dopamine release then activates pup LG in lactating female rats (Champagne, Stevenson, Gratton, & Meaney, 2004; Shahrohk et al., in press). This effect is abolished with the infusion of an oxytocin receptor antagonist into the ventral tegmental area (Shahrohk et al., in press). Likewise, the same oxytocin receptor antagonist completely eliminates the differences in maternal behavior between high- and low-LG mothers (Champagne et al., 2001). Drugs that block the reuptake of synaptic dopamine, thus increasing the overall dopaminergic signal, increase pup LG in low-LG mothers and eliminate the difference in maternal behavior between high- and low-LG mothers. In summary, the increased estrogen receptor a expression in the MPOA of high-LG mothers leads to greater sensitivity to estrogen, an increased level of oxytocin levels in MPOA neurons that project directly to the dopamine neurons in the ventral tegemental regions, and greater activation of dopamine release during nursing bouts. And the differences in estrogen receptor a expression, like those in pup LG, are reversed with cross-fostering (Champagne, Francis, et al., 2003), suggesting that maternal care regulates the activity of the estrogen receptor in the MPOA, which then forms the basis for subsequent "inherited" differences in maternal behavior.

This finding reveals another example of a maternal effect on gene expression, and there is evidence for epigenetic mediation. The activation of the estrogen receptor α-gene in the rat brain occurs through the estrogen receptor 1B promoter. This promoter contains multiple cytosine sites that are potential targets for DNA methylation. Champagne et al. (2006) found increased cytosine methylation across the exon 1B promoter in the offspring of low-LG mothers. Activation of the exon 1B promoter occurs in response to the binding of the transcription factor Stat5, and the adult offspring of low-LG mothers show decreased Stat5 association with the exon 1B promoter in the MPOA. These findings suggest that differences in DNA methylation may mediate the effect of maternal care on the expression of estrogen receptor α in the MPOA and thus serve as the molecular basis for the nongenomic transmission of individual differences in maternal behavior from the mother to her female offspring.

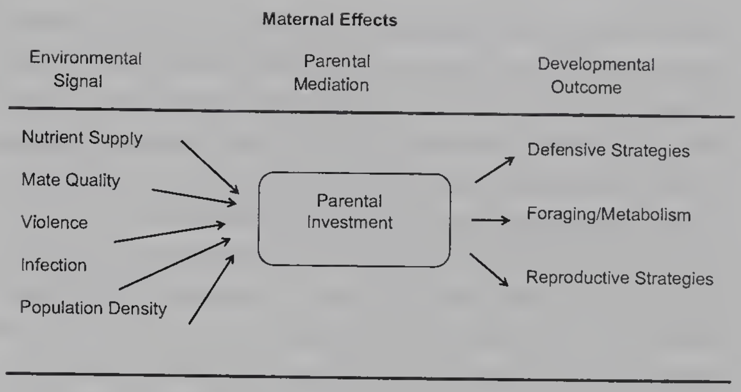


Figure 6. Summary of the literature from evolutionary biology on "maternal effects."

Note. The consistent theme reflected in these studies is that various environmental signals can alter multiple phenotypic outcomes through effects on parent–offspring interactions, broadly referred to as *parental investment*. The relevant form of the variation in parent–offspring interaction (investment) will vary depending upon the species. The principal idea is that of parental mediation and of coordinated effects on multiple phenotypic outcomes.

Adaptive Value of Epigenetic Maternal Effects

Variations in parental signals alter gene expression and thus the development of individual differences in complex phenotypes. But why bother? Why would nature configure such a process? Why transmit individual differences in stress reactivity or maternal behavior across generations through a process that is driven by parental care and mediated by the complex cellular machinery necessary to rearrange epigenetic markings on the DNA? Why not simply leave such issues of inheritance in the hands of classic genetic transmission? The answer may lie in the simple fact that unlike nucleotide sequence, epigenetic marks are dynamic and indeed reversible. The more flexible epigenetic mechanism would provide the basis for an adaptive parental effect: In the rat at least, parents can actively remodel epigenetic marks and thus affect patterns of gene expression in the offspring. Such effects do not occur at the level of nucleotide sequence.

Studies in the fields of evolutionary biology and ecology report maternal effects on phenotype in the offspring across a wide range of species (see Figure 6; Badyaev, 2008; Cameron et al., 2005; Meaney, 2007; Mousseau & Fox, 1998; Rossiter, 1998). Despite the fact that these studies have been performed largely with simpler organisms, such as plants, insects, and reptiles, the emerging theme is much the same as that described earlier in studies with human families: Environmental adversity decreases parental investment in the offspring and thus alters phenotypic development. And there is evidence that these phenotypic effects are adaptive within adverse settings (Mousseau & Fox, 1998).

Obviously, the form of "parental investment" varies across species. In plants, the variation may involve seed quality. Among certain insects "investment" would include the nutrient value of the "propulgate," the food source left by the mother for the hatching offspring. While some insects do exercise postnatal behavioral care of the offspring, variations in this form of investment are most relevant for mammalian species. Nevertheless, the stability of the relation between the quality of the prevailing environment, parental investment, and phenotypic development in the offspring led Hinde (1986) to suggest that evolution may have actually shaped the offspring to "use" parental signals, such as variations in parental care behaviors, to forecast the quality of the environment in which they must function (also see Bateson, 1994; Bateson et al., 2004; Rossiter, 1998). Evolution should come to favor offspring that are able to accurately "read" variations in parental behavior as forecasts of environmental conditions and thus as useful signposts to guide developmental outcomes (Hinde, 1986). By definition, such responses should occur in reaction to variation within the normal range. Why evolve responses to forms of parental care that are unlikely to occur? Parents (or parent in some cases) are a logical source of such information since they are the one "environmentally informed" and constant experience for the offspring. Moreover, since parents are invested in the adaptive success of their biological offspring, one would expect that the fidelity of signals emanating from a parent would be greater than that of other adult conspecifics. Thus, to the extent that the parent and offspring share a common interest in the adaptive value of such phenotypic plasticity, selection may also act on the signaling capacity of the parent. Indeed, phenotypic plasticity in response to parental signals may also be thought of as a parental strategy, although it is clear that the benefits work in both directions as would be expected among organisms that share genes. In either case, the sensitivity of the offspring to parental signals during critical phases of development may be thought of as a strategy that favors a highly predictable relation between environmental conditions, parental input, and phenotypic variation in defensive responses. For the offspring, parents should matter.

Hinde's (1986) formulation implies an adaptive function for parental effects in preparing the offspring for the environmental demands that are likely to prevail during development and into adulthood (also see Bateson et al., 2004; Gluckman & Hanson, 2004; Meaney, 2001). Indeed, it is probably true that the environmental conditions of the adult are a reasonable predictor of those facing the offspring (with modern, urbanized human populations as a rare exception; Gluckman & Hanson, 2004; West & King, 1987). However, environmental conditions can vary with the migration of populations. Moreover, even within a similar environment, the ability of individuals to access resources can vary from one individual to another as a function of, for example, social dominance hierarchies. Success requires the ability to adapt, to fine-tune phenotypic development in relation to the prevailing environmental demands. Learning, of course, would provide a mechanism for such phenotypic plasticity. But learning requires trial and, unfortunately, error. Nature does not always permit such indulgences. Epigenetic variation may provide a mechanism for the dynamic regulation of the genomic machinery in response to variation in prevailing environmental conditions during development (Jablonka & Lamb, 2005). Such epigenetic adaptations could then occur directly as a result of the relevant environment condition (e.g., nutrient availability) or indirectly through parental mediation. Such effects could, for example, explain the enduring effects of childhood socioeconomic status on health outcomes in humans.

We suggest that parental signals over the perinatal period serve as an important catalyst for epigenetic remodeling of the genome. If such effects support an adaptive function, then there should be a high degree of fidelity between the quality of the environment and parental care. This argument rests upon at least three assumptions. First, there is a consistent relation between the quality of the environment and that of parental care. Second, variations in specific developmental outcomes should be adaptive. This includes an adaptive value to traits that appear to increase vulnerability to illness, such as increased stress reactivity. And third, the ecology of the neonate reliably predicts the conditions of adulthood.

Stress and Parenting

Perhaps, the most compelling evidence for a direct effect of environmental adversity on parent–offspring interactions emerges from the studies of Rosenblum, Coplan, and colleagues with nonhuman primates (Coplan, Andrews, Rosenblum, & Nemeroff, 1996). Bonnet macaque mother–infant dyads were maintained under typical lab conditions, with free, unhindered access to food (a low-foraging-demand condition) or one in which the amount of available food varied and required long periods of searching (high-foraging-demand condition). The high-foraging-demand condition severely disrupted mother–infant interactions, producing significant conflict. Infants of mothers housed under these conditions were more timid and fearful, with some revealing signs of depression that are commonly observed in maternally separated macaque infants. Such reactions were even apparent among infants that were in contact with their mothers. As adolescents, the infants reared in the high-foraging-demand conditions were more fearful, submissive, and showed less social play behavior. As expected, these conditions also affected the development of neural systems that mediate behavioral and endocrine response to stress. Adult monkeys reared under variable foraging demand (VFD) conditions showed increased central levels of CRF and increased noradrenergic responses to stress. It would be fascinating to see if such traits would then be transmitted to the next generation. In some rather remarkable cross-fostering studies, Maestripieri (2005) has shown that individual differences in maternal behavior are transmitted across generations in the rhesus monkey.

The critical issue here is that of a direct effect of environmental adversity on maternal behavior. Stress during pregnancy decreases maternal responsivity in lactating rats (Fride, Dan, Gavish, & Weinstock, 1985; Kinsley, Mann, & Bridges, 1998; Moore & Power, 1986). Gestational stress also eliminates the differences in pup LG between high- and low-LG mothers (Champagne & Meaney, 2006). Gestational stress decreases the frequency of maternal LG in the high-, but not in low-LG mothers (also see Smith, Seckl, Evans, Costall, & Smythe, 2004), and the effect is apparent even with subsequent litters in the absence of further exposure to stress. As expected, the effects on maternal behavior are apparent in the development of the offspring. As adults, the offspring of high-LG/gestationally stressed mothers were comparable to those of low-LG dams on measures of maternal behavior, as well as in fear behavior and hippocampal glucocorticoid receptor gene expression. These effects can be distinguished from those associated with prenatal stress in animals that were in utero during the imposition of the stressor by simply examining the offspring of subsequent litters in which the behavior of the mother remains affected, but in the absence of the stressor (Champagne & Meaney, 2006).

As mentioned earlier, there is considerable, albeit correlational, evidence for a relation between environmental adversity and parental care in humans. Environmental adversity influences emotional well-being in parents and these effects are reflected in alterations in parental care, commonly reflecting a decreased level of investment (Fleming, 1988; Repetti et al., 2002). Increased maternal stress is associated with less sensitive child care (Dix, 1991; Goldstein, Diener, & Mangelsdorf, 1996). The children of highly stressed primary caregivers tend to develop insecure parental attachment (Goldstein et al., 1996; Vaughn, Egeland, Sroufe, & Waters, 1979). Parents in conditions of poverty experience more negative emotions, irritability, depressed, and anxious moods, which lead to more punitive parenting (Belsky, 1993; Conger, McCarty, Yang, Lahey, & Kropp, 1984; Grolnick, Gurland, DeCourcey, & Jacob, 2002). The resulting patterns of parental behavior can affect the development of HPA responses to stress in the offspring (Heim et al., 2000; Pruessner, Champagne, Meaney, & Dagher, 2004), an effect that may associate with altered glucocorticoid receptor expression (McGowan et al., 2009). Hane and Fox (2006) found a direct relation between the quality of maternal care and behavioral inhibition in children. There is rather compelling evidence for parental mediation in the effects of impoverished environments on phenotypic development. Thus, the effects of poverty on emotional and cognitive development are, in part, mediated by variations in parent–offspring interactions. If parental care factors are statistically controlled, there no longer remains any discernible effect of poverty of child development (Conger, Ge, Elder, Lorenz, & Simons, 1994; McLoyd, 1998). Moreover, parents and parenting style are highly effective targets for intervention studies aimed at development outcomes in children living in adverse environmental conditions (Belsky, 1997; Fisher et al., 2000; Offord et al., 1992; Olds et al., 1998; Van den Boom, 1994).

Adaptive Value of Increased Stress Responses

Intergenerational transmission of individual differences in stress reactivity via parental behavior could represent an adaptive approach to development. Since the offspring usually inhabit a niche that is similar to their parents, the transmission of individual differences in traits from parent to offspring could serve to be adaptive with respect to survival. The conceptually challenging feature of this argument is that it requires that we identify an adaptive virtue to a phenotypic profile, such as increased HPA stress responses, that predict greater vulnerability to chronic illness.

Adverse environmental conditions such as those associated with poverty over the adult life of the parent have historically predicted more of the same for the offspring. The critical challenge over the course of development is to mold specific features of phenotype in a manner that is most appropriate to the level of environmental demand. This capacity for phenotypic plasticity (Agrawal, 2001; Mousseau & Fox, 1998) is a product of evolutionary forces. As such, it is critical to note that the ultimate measure of success for any phenotypic trait is the degree to which it enhances the probability for reproductive success. The health of the individual is a relevant consideration *only to the degree that it influences the ability to successfully reproduce.* Health will influence survival and the ability to attract mating opportunities, and may therefore be of relevance, at least during the period of active reproduction. However, the incidence of chronic illnesses that commonly occur in later life, once the prospects for reproductive activity have declined, is not a major consideration in evaluating the adaptive value of any specific trait in the evolutionary sense.

There is a very real divide between the way in which the success of development is evaluated by those working in the health sciences as compared to those in biology. Optimal development for those in the health sciences is judged by the quality of life and the absence of disease. For biologists, success is measured in the terms of the currency of natural selection—reproductive fitness. For biologists, there are no ideal phenotypes. Rather, the adaptive merits of phenotype are apparent only in relation to success within a particular set of environmental conditions: One single phenotype does not fit all. For example, an increase in the responsivity of an individual to threat (i.e., greater stress reactivity) is realistically considered as a risk factor for multiple forms of chronic illness. Thus, the catabolic effects associated with adrenal glucocorticoids or sympathetic catecholamines tend to be vilified, especially in psychology and psychiatry. Indeed, chronic elevations in the levels of these "stress mediators" can directly promote illness (Chrousos & Gold, 1992; Dallman et al., 2005; McEwen, 2007; Sapolsky et al., 2000; Walker, 2007). However, the bitter truth of adaptation and survival in the face adversity is that such hormonal stress responses are essential for continued life. The survival interests of an individual during periods of increased environmental demand are well served by behavioral (e.g., vigilance, fearfulness) and endocrine (HPA and metabolic/cardiovascular) responses to stress. These responses promote detection of potential threat, fear conditioning to stimuli associated with threat and avoidance learning. Moreover, the hormonal effectors of sympathoadrenal and HPA stress responses mobilize energy reserves through effects of lipolysis, glycolysis, and gluconeogenesis. These effects are the hallmark of the shift to catabolism that occurs during periods of stress and are essential for animals exposed to chronic stress, particularly if the stressor is coupled to conditions of famine. Indeed, the ability to survive sustained periods of nutrient deprivation depends upon the capacity to increase circulating levels of glucocorticoids and catecholamines. Impoverished environments are also commonly associated with multiple sources of infection. Under such conditions adrenal glucocorticoids serve as a potent defense against septic shock (Munck et al., 1984; Sapolsky et al., 2000). Among rats, animals with increased HPA responses to agents such as bacterial endotoxins are at reduced risk for sepsis. Interestingly, adults exposed to a bacterial endotoxin during the 1st week of life exhibit increased HPA responses to stress as well increased resistance to sepsis upon subsequent exposure to bacterial infection (Shanks, Larocque, & Meaney, 1995; Shanks et al., 2000). Conversely, postnatal conditions that increase maternal LG and dampen HPA responses to stress *increase* vulnerability to endotoxin-induced sepsis. These findings underscore the potentially adaptive value of increased HPA and sympathetic responses to stress, especially for individuals living under conditions of impoverishment and infection. Thus, the decreased parental investment associated with more stressful conditions may in fact be of benefit to the offspring, if indeed such conditions remain stable over time. Decreased parental investment might enhance the stress responses of the offspring, which could serve to protect against repeated periods of nutrient deprivation and infection. This theme is apparent in the effects of maternal care on neurocognitive development in the rat.

Maternal Care and Neurocognitive Development

Pup LG in the rat dynamically alters the activity of endocrine systems that favor of somatic growth (Levine, 1994; Schanberg et al., 1984). Comparable effects are apparent in the hippocampus, a brain region intimately associated with learning and memory. The neonatal offspring of high-LG mothers show increased hippocampal expression of genes that encode for neurotrophic factors, which supports the sprouting and survival of synapses, this effect includes the expression glutamate receptor subunits. The activation of glutamate receptors in the hippocampus stimulates synaptogenesis (Kirkwood, Dudek, Gold, Aizenman, & Bear, 1993; Schatz, 1990). Thus, as adults, the offspring of high-LG mothers show increased hippocampal synaptic density (Bredy, Humpartzoomian, Cain, & Meaney, 2003; Champagne et al., 2008; Liu et al., 2000) associated with more extensive dendritic arborization of both pyramidal and granule cell populations (Bagot et al., 2009; Champagne et al., 2008). The effect of maternal care on synaptic development suggests an increased capacity for synaptic plasticity in the offspring of high-LG mothers. Long-term potentiation (LTP) is an electrophysiological model of synaptic plasticity at specific neural sites that is thought to mimic the structural remodeling of synaptic connections that underlie learning and memory. Hippocampal LTP formation is indeed stronger in the adult offspring of High compared to Low LG mothers (Bagot et al., 2009; Bredy et al., 2003; Champagne et al., 2008). And not surprisingly, the adult offspring of high-LG mothers show improved performance hippocampal-dependent tests of learning and memory, such as spatial learning (Bredy, Zhang, Grant, Diorio, & Meaney, 2004; Liu et al., 2000) and object recognition (Bredy et al., 2004).

These findings suggest a rather predictable effect of maternal care on hippocampal development, involving enhanced synaptic development in neonatal life and an increased capacity for synaptic plasticity in adulthood, revealed in electrophysiological as well as behavioral studies. Such findings would seem to reflect an advantage for the offspring of high-LG mothers. But these findings fail to consider the issue of context. Studies of in vitro LTP are performed with hippocampal slices obtained from animals in the resting state. Studies of in vivo LTP or behavioral tests of learning and memory are performed in animals habituated to the testing conditions. What if animals were tested under more demanding, stressful conditions? Champagne et al. (2008) and Bagot et al. (2009)

examined LTP in hippocampal slices obtained from the adult offspring of high- or low-LG mothers under basal conditions, or in the presence of stress-like levels of glucocorticoids. Glucocorticoids are an endocrine signature of the stress response and known to diminish hippocampal LTP (Bodnoff et al., 1995; Diamond & Rose, 1994; Joels, Karst, Krugers, & Lucassen, 2007; Kim & Yoon, 1998; Shors, Foy, Levine, & Thompson, 1990). This effect is readily apparent in sections obtained from high-LG mothers. LTP under basal conditions is markedly greater in hippocampal slices from high- compared to low-LG offspring, and the magnitude of LTP in slices from high-LG offspring is markedly reduced by glucocorticoids. The exact *opposite* is observed in hippocampal slices prepared from the offspring of low-LG mothers: Glucocorticoids significantly enhance LTP to the levels observed in hippocampal slices from high-LG mothers under basal conditions. This is true for LTP obtained from multiple regions within the hippocampus (i.e., for aficionados of hippocampal anatomy, this included neurons within both the Ammon's Horn and the dentate gyrus). And an analogous effect is apparent at the level of learning and memory. When the adult offspring of high- and low-LG mothers are tested not under the benign conditions described earlier but in tests on contextual fear conditioning, the performance of the offspring of the low-LG mothers is significantly improved (i.e., increased learning of the context–shock contingency) over that of high-LG dams (Bagot et al., 2009; Champagne et al., 2008). Contextual fear conditioning is also a hippocampal-dependent test of learning and memory (Fanselow, 2000; Maren & Quirk, 2004), in this case one that involves the association of a context with an aversive event (i.e., a mild shock). Thus, both hippocampal synaptic plasticity and hippocampal-dependent learning and memory are enhanced in the offspring of low-LG mothers *under stressful conditions*. The effects of maternal care on hippocampal function are context dependent.

Implications

All cellular processes derive from a constant dialogue between the genome and environmental signals. Thus, genotype–phenotype relations are defined by the context within which the genome operates. Likewise, the consequences for Gene × Environment interactions at the level of function are defined by the broader context, including the demands of the prevailing environment. The results of studies on the effects of maternal care on hippocampal development in the rat reflect an important point. Although it is tempting to assume that increased pup LG enhances synaptic plasticity and cognitive performance, such effects are apparent only under conditions that are minimally stressful. Under stressful conditions, such as those typified by contextual fear conditioning, it is the offspring of low-LG mothers that show improved synaptic plasticity and learning. We suggest that such findings represent a situation in which parental care shapes adaptive phenotypes (Hinde, 1986; Meaney, 2001; Bateson et al., 2004; Zhang et al., 2004; Gluckman & Hanson, 2007). In the rat, environmental stressors decrease parental investment and enhance behavioral and endocrine responses to stress, as well as learning and memory under stressful conditions. If indeed the decreased parental investment accurately reflects an increased level of environmental demand for the offspring, then such effects could be highly adaptive. Regardless of the merits of this proposal, these findings attest to the simple fact that the adaptive value of any phenotypic profile depends upon the environmental context: There is no universally "ideal" phenotype.

There is also evidence for the potentially adaptive effects of increased stress reactivity that more directly bears on the interests of child and adolescent psychology. The research of Farrington, Gallagher, Morley, St Ledger, and West (1988) and Tremblay (e.g., Haapasap & Tremblay, 1994) on young males growing up in impoverished, high-crime urban settings illustrates the potential advantages of increased emotional stress reactivity. Both studies show that shy and more timid males are most successful in avoiding the pitfalls associated with "criminogenic" environments. Under such conditions, behavioral inhibition emerges as a protective factor (Haapasap & Tremblay, 1994), despite the fact that this same profile is associated with an increased risk for mood disorders in later life (Pérez-Edgar & Fox, 2005). Moreover, under such adverse conditions a parental rearing style that favored the development of a greater level of stress reactivity to threat could be viewed as adaptive. If indeed there is no single ideal phenotype, then it should follow *that there is no single ideal form of parenting*. If this conclusion has worth, then it leads us to question the wisdom of establishing parenting programs that foster parental skills based on studies of families rearing children under more favorable conditions.

Caveats

The magnitude and scope of the maternal effects on rodent development should not be surprising. Hinde (1986), Bateson (1994), and others have described the logic for the importance of maternal signals for the developing mammal. Indeed, even the gametes include parental signals apart from genomic DNA, such as RNAs and proteins, which influence embryonic development. Nevertheless, it is important to understand that the rodent studies described earlier are designed to examine how variations in maternal *can* affect development; they do not address the issue of *how* and to *what degree* maternal care influences development under normal circumstances in the rat. Thus, the model described earlier examines maternal effects on neural development in animals that are then housed under highly standardized conditions following the weaning period, eliminating the possibility that maternal effects might be subsequently altered by variations in the postweaning environment. Indeed, there is evidence that postweaning environments can reverse the effects of preweaning variations in maternal care in the rat (Bredy et al., 2003; Champagne et al., 2006). These studies are perhaps best considered as a model for the mechanisms by which social experience in early life can influence the structure and operation of the genome with respect to specific phenotypic outcomes. Indeed, as suggested by Belsky (1997), Rutter (2007), and others, the effects of environmental influences, such as maternal care, will vary across individuals, and the studies of Fox et al. (2005) in humans and Suomi and colleagues (Bennett et al., 2002; Champoux et al., 2002; Suomi, 2006) in the nonhuman primate, suggest that variation in genomic sequence may be a critical factor in determining sensitivity to parental signals.

The study of epigenetics provides a molecular mechanism for environmental effects of gene expression. But the ultimate effect on gene transcription is best thought of as an emergent property of the interaction between the epigenetic state and the underlying genome. Thus, genomic sequence determines the effects of epigenetic states on gene transcription. While DNA methylation is linked to transcriptional silencing, such relations are not universal and the strength of the relation may be determined by the underlying genomic sequence (Weber et al., 2007). For example, among promoters with a lower percentage of CpG dinucleotides, there appears to be a weaker relation between the overall level of methylation and that of gene transcription. Likewise, the presence of histone modifications (H3K4me) that associate with transcriptionally active genes appears to be determined not only by DNA methylation but also by the underlying genomic sequence. Such findings should dissuade us from assuming that genomic sequence is merely passive player in the definition of epigenetic states. The underlying genomic sequence influences the nature of potential epigenetic states as well as their importance for gene transcription. Finally, many transcriptionally inactive genes show relatively unmethylated promoters. While this finding does not preclude the possibility that specific modifications in the methylation status at critical CpG sites may define changes in gene transcription, such as occurs for the *IL2* gene (Murayama et al., 2006), it is important to bear in mind that DNA methylation is only one level of influence in the very complex biochemical machinery that regulates gene transcription. These are still early days in the study of the molecular mechanisms for the environmental programming of gene expression. But the door has clearly been opened.

General Conclusions

The effects of maternal care on gene expression and neural function in the rat provide an understanding of how environmental events, including variations in parent–offspring interactions at the level of behavior, can become physically imprinted upon the genome. Maternal care can directly alter intracellular signals that, in turn, structurally alter the DNA and its operation. These structural modifications involve DNA methylation, a classic epigenetic mark that regulates gene transcription. More recent studies suggest that comparable epigenetic modifications associate with learning and memory (Lubin et al., 2008; Miller & Sweatt, 2007), chronic exposure to drugs of abuse (Renthal & Nestler, 2008), and psychiatric illness (Grayson et al., 2005; Ptak & Petronis, 2008; Tsankova et al., 2007). The dynamic genome is probably a slightly foreign concept to those who imagine the DNA as simply the repository of the sequence information that forms what is commonly referred to as the "genetic code." But the research of the postgenomic era, with its focus on the operation rather than simply the composition of the genome, reveals that the DNA is an active target for remodeling by cellular signals that are activated by environmental events. The reality of the functional genome does not admit to main effects of either gene or environment, but rather to a constant interaction between the DNA and its environment.

The study of epigenetic mechanisms reveals the importance of DNA remodeling for gene transcription, which is certainly the most fundamental operation of the genome. The transcription of the genome requires dynamic alterations to the chromatin structures within which the DNA operates. Such events involve transient modifications of the histone proteins that determine the accessibility of the DNA. Indeed, cellular signals that increase or decrease gene transcription often function within complexes that include enzymes that directly operate on chromatin. Potentially more stable signals, such as DNA methylation, are also regulated by environmental events and regulate chromatin activity over longer periods. Such epigenetic marks are thus a candidate mechanism for the environmental programming of gene expression. This level of complexity is an essential feature of biology. The ability of an organism to adapt to variations in environmental conditions over the course of its lifespan demands plasticity in genotype–phenotype relations.

The emergence of an integrative developmental perspective has been constrained by historical misunderstandings of the processes by which variations in genomic sequence and nongenomic factors contribute to the development of individual differences in any specific traits. While quantitative behavioral genetics is a valid approach in establishing patterns of familial transmission (i.e., heritability) that *might* imply a causal role for genomic variation, such studies are commonly subject to misinterpretation of genomic influences. The heritability of individual differences could result as a function of nongenomic biological signals such as RNA and protein emanating from the parent as well as from instances where epigenetic modifications enter the germline (Anway, Cupp, Uzumcu, & Skinner, 2005; Jablonka & Lamb, 2005; Vardhman et al., 2003; Whitelaw & Whitelaw, 2006). There are multiple processes by which individual differences in complex traits might be transmitted from parent to offspring. Moreover, estimates of the contribution of *either* genetic or environmental influences at the level of individual variation are complicated by the reality of Gene × Environment interactions (Lewontin, 1974). Most troublesome is the obvious disjunction between the definition of genomic influences that derives from statistical models and the actual biological reality of the genome. It is the operation of the genome that directly influences phenotype. The operation of the genome at any phase of the life cycle is an emergent property of the constant and very physical interaction of the genome with environmentally regulated, intracellular signals that directly alter chromatin structure. Thus, function at any level of biology emerges as a function of the continuous dialogue between the genome and its environment. Attempts to parse the influence of genomic and environmental influences on the expression of complex traits are inconsistent with even the most rudimentary understanding of gene function.

These issues are critical for the study of child development. Research in this field guides both the creation and the evaluation of intervention studies (e.g., Fisher et al., 2000; Olds et al., 1998). The science and technology of the genomic era offer remarkable opportunities to enhance the sophistication of our approach to these challenges. We are now positioned to study directly the origins of individual differences in sensitivity/resistance to treatment within the context of intervention studies (Belsky, 1997; Rutter, 2007). Indeed, the many years of remarkable research in child psychology and psychiatry forms the basis for hypothesis-driven studies of candidate Gene × Environment interactions of responses to treatment within the context of intervention studies. Thus, the response of a child to psychosocial interventions, such as those targeting mother–infant attachment, is influenced by the temperament or reactivity of the child (Belsky, 1997). More reactive children also seem to be more sensitive to the quality of the rearing environment (Boyce & Ellis, 2005). The temperament of the child is influenced by the presence of specific genomic variations in genes that encode for proteins involved in the function of the 5-HT and dopamine systems (Auerbach et al., 2002; Ebstein, 2003; Fox et al., 2005; Lakatos et al., 2003). Could such genomic variations mediate the observed effect of child temperament on the response to specific interventions? What might be a more appropriate intervention for "treatment-resistant" populations? A more integrated Gene × Environment approach offers the opportunity to understand clearly the nature of individual differences in vulnerability or resistance for psychopathology and to more effectively target interventions.

References

Agrawal, A. A. (2001). Phenotypic plasticity in the interactions and evolution of species. *Science, 294,* 321–326.

Akbarian, S., & Huang, H. S. (2009). Epigenetic regulation in human brain-focus on histone lysine methylation. *Biological Psychiatry, 65,* 198–203.

Anway, M. D., Cupp, A. S., Uzumcu, M., & Skinner, M. K. (2005). Epigenetic transgenerational actions of endocrine disruptors and mate fertility. *Science, 308*, 1466–1469.

Auerbach, J., Geller, V., Lezer, S., Shinwell, E., Belmaker, R. H., Levine, J., et al. (1999). Dopamine D4 receptor (D4DR) and serotonin transporter promoter (5-HTTLPR) polymorphisms in the determination of temperament in 2-month-old infants. *Molecular Psychiatry, 4*, 369–373.

Auerbach, J., Geller, V., Lezer, S., Shinwell, E., Belmaker, R. H., Levine, J., et al. (2002). Dopamine D4 receptor (D4DR) and serotonin transporter promoter (5-HTTLPR) polymorphisms in the determination of temperament in 2-month-old infants. *Molecular Psychiatry, 4*, 369–373.

Badyaev, A. V. (2008). Maternal effects as generators of evolutionary change. *Annals of the New York Academy of Sciences, 1133*, 151–161.

Bagot, R. C., van Hasselt, F. N., Champagne, D. L., Meaney, M. J., Krugers, H. J., & Joëls, M. (2009). Maternal care determines rapid effects of stress mediators on synaptic plasticity in adult rat hippocampal dentate gyrus. *Neurobiology of Learning & Memory, 92*, 292–300.

Bakermans-Kranenburg, M. J., Van IJzendoorn, M. H., Pijlman, F. T., Mesman, J., & Juffer, F. (2008). Experimental evidence for differential susceptibility: Dopamine D4 receptor polymorphism (DRD4 VNTR) moderates intervention effects on toddlers' externalizing behavior in a randomized controlled trial. *Developmental Psychology, 44*, 293–300.

Bale, T. L., & Vale, W. W. (2004). CRF and CRF receptors: Role in stress responsivity and other behaviors. *Annual Review of Pharmacology & Toxicology, 44*, 525–557.

Bateson, P. (1994). The dynamics of parent–offspring relationships in mammals. *Trends in Ecology and Evolution, 9*, 399–403.

Bateson, P., Barker, D., Clutton-Brock, T., Deb, D., D'Udine, B., Foley, R. A., et al. (2004). Developmental plasticity and human health. *Nature, 430*, 419–421.

Belsky, J. (1993). Etiology of child maltreatment: A developmental-ecological analysis. *Psychological Bulletin, 114*, 413–434.

Belsky, J. (1997). Theory testing, effect-size evaluation, and differential susceptibility to rearing influence: The case of mothering and attachment. *Child Development, 64*, 598–600.

Bennett, A. J., Lesch, K.-P., Heils, A., Long, J. C., Lorenz, J. G., Shoaf, S. E., et al. (2002). Early experience and serotonin transporter gene variation interact to influence primate CNS function. *Molecular Psychiatry, 7*, 118–122.

Bestor, T. H. (1998). Gene silencing. Methylation meets acetylation. *Nature, 393*, 311–312.

Bettegowda, A., & Smith, G. W. (2007). Mechanisms of maternal mRNA regulation: Implications for mammalian early embryonic development. *Frontiers in Bioscience, 12*, 3713–3726.

Bird, A. P. (1986). CpG-rich islands and the function of DNA methylation. *Nature, 321*, 209–213.

Bird, A. P. (2002). DNA methylation patterns and epigenetic memory. *Genes Development, 16*, 6–21.

Bird, A. (2007). Perceptions of epigenetics. *Nature, 447*, 396–398.

Bird, A. P., & Wolffe, A. P. (1999). Methylation-induced repression—Belts, braces, and chromatin. *Cell, 99*, 451–454.

Blier, P., & de Montigny, C. (1999). Serotonin and drug-induced therapeutic responses in major depression, obsessive-compulsive and panic disorders. *Neuropsychopharmacology, 21*, 91S–98S.

Bodnoff, S. R., Humphreys, A. G., Lehman, J. C., Diamond, D. M., Rose, G. M., & Meaney, M. J. (1995). Enduring effects of chronic corticosterone treatment on spatial learning, synaptic plasticity, and hippocampal neuropathology in young and mid-aged rats. *Journal of Neuroscience, 15*, 61–69.

Boulinier, T., & Staszewski, V. (2008). Maternal transfer of antibodies; raising immuno-ecology issues. *Trends in Ecology and Evolution, 23*, 282–288.

Boyce, W. T, & Ellis, B. J. (2005). Biological sensitivity to context: 1. An evolutionary-developmental theory of the origins and functions of stress reactivity. *Development and Psychopathology, 17*, 271–301.

Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Nemes, A., et al. (1994). Sp1 elements protect a CpG island from de novo methylation. *Nature, 371*, 435–438.

Bredy, T. W., Humpartzoomian, R. A., Cain, D. P., & Meaney, M. J. (2003). The influence of maternal care and environmental enrichment on hippocampal development and function in the rat. *Neuroscience, 118*, 571–576.

Bredy, T. W., Wu, H., Crego, C., Zellhoefer, J., Sun, Y. E., & Barad, M. (2007). Histone modifications around individual BDNF gene promoters in prefrontal cortex are associated with extinction of conditioned fear. *Learning & Memory, 14*, 268–276.

Bredy, T. W., Zhang, T.-Y., Grant, R. J., Diorio, J., & Meaney, M. J. (2004). Peripubertal environmental enrichment reverses the effects of maternal care on hippocampal development and glutamate receptor subunit expression. *European Journal of Neuroscience, 20*, 1355–1362.

Bronnegard, M., & Okret, S. (1991). Regulation of the glucocorticoid receptor in fetal rat lung during development. *Journal of Steroid Biochemistry & Molecular Biology, 39*, 13–17.

Bruniquel, D., & Schwartz, R. H. (2003). Selective, stable demethylation of the interleukin-2 gene enhances transcription by an active process. *Nature Immunology, 4*, 235–240.

Burton, C. L., Chatterjee, D., Chatterjee-Chakraborty, M., Lovic, V., Grella, S. L., Steiner, M., et al. (2007). Prenatal restraint stress and motherless rearing disrupts expression of plasticity markers and stress-induced corticosterone release in adult female Sprague–Dawley rats. *Brain Research, 1158*, 28–38.

Caldji, C., Diorio, J., & Meaney, M. J. (2003). Variations in maternal care alter GABAA receptor subunit expression in brain regions associated with fear. *Neuropsychopharmacology, 28*, 150–159.

Caldji, C., Francis, D. D., Sharma, S., Plotsky, P. M., & Meaney, M. J. (2000). The effects of early rearing environment on the development of GABAA and central benzodiazepine receptor levels and novelty-induced fearfulness in the rat. *Neuropsychopharmacology, 22*, 219–229.

Caldji, C., Tannenbaum, B., Sharma, S., Francis, D. D., Plotsky, P. M., & Meaney, M. J. (1998). Maternal care during infancy regulates the development of neural systems mediating the expression of behavioral fearfulness in adulthood in the rat. *Proceedings of the National Academy of Sciences, USA, 95*, 5335–5340.

Cameron, N., Parent, C., Champagne, F. A., Fish, E., Ozaki-Kuroda, K., & Meaney, M. J. (2005). The programming of individual differences in defensive responses and reproductive strategies in the rat through variations in maternal care. *Neuroscience & Biobehavioral Reviews, 29*, 843–865.

Canli, T., & Lesch, K.-P. (2007). Long story short: The serotonin transporter in emotion regulation and social cognition. *Nature Neuroscience, 10*, 1103–1109.

Canli, T., Qiu, M., Omura, K., Congdon, E., Haas, B. W., Amin, Z., et al. (2006). Neural correlates of epigenesis. *Proceedings of the National Academy of Sciences, USA, 103*, 16033–16038.

Carola, V., Frazzetto, G., Pascucci, T., Audero, E., Puglisi-Allegra, S., Cabib, S., et al. (2008). Identifying molecular substrates in a mouse model of the serotonin transporter × environment risk factor for anxiety and depression. *Biological Psychiatry, 63*, 840–846.

Caspi, A., & Moffit, T. E. (2006). Gene–environment interactions in psychiatry: Joining forces with neuroscience. *Nature Reviews Neuroscience, 7*, 583–590.

Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., et al. (2003). Influence of life stress on depression: Moderation by a polymorphism in the 5-HTT gene. *Science, 301*, 386–389.

Champagne, F. A. (2008). Epigenetic mechanisms and the transgenerational effects of maternal care. *Frontiers in Neuroendocrinology, 29*, 386–397.

Champagne, F. A., Diorio, J., Sharma, S., & Meaney, M. J. (2001). Variations in maternal care in the rat are associated with differences in estrogen-related changes in oxytocin receptor levels. *Proceedings of the National Academy of Sciences, USA, 98*, 12736–12741.

Champagne, F. A., Francis, D. D., Mar, A., & Meaney, M. J. (2003). Naturally-occurring variations in maternal care in the rat as a mediating influence for the effects of environment on the development of individual differences in stress reactivity. *Physiology & Behavior, 79*, 359–371.

Champagne, F. A., & Meaney, M. J. (2006). Stress during gestation alters postpartum maternal care and the development of the offspring in a rodent model. *Biological Psychiatry, 59*, 1227–1235.

Champagne, F. A., Stevenson, C., Gratton, A., & Meaney, M. J. (2004). Individual differences in maternal behavior are mediated by dopamine release in the nucleus accumbens. *Journal of Neuroscience, 24*, 4113–4123.

Champagne, D. L., van Hasselt, F., Ramakers, G., Meaney, M. J., de Kloet, E. R., Joels, M., et al. (2008). Maternal care alters dendritic length, spine density and synaptic potentiation in adulthood. *Journal of Neuroscience, 28*, 6037–6045.

Champagne, F. A., Weaver, I. C. G., Diorio, J., Dymov, S., Szyf, M., & Meaney, M. J. (2006). Maternal care regulates methylation of the estrogen receptor alpha 1b promoter and estrogen receptor alpha expression in the medial preoptic area of female offspring. *Endocrinology, 147*, 2909–2915.

Champagne, F. A., Weaver, I. C. G., Diorio, J., Sharma, S., & Meaney, M. J. (2003). Natural variations in maternal care are associated with estrogen receptor alpha expression and estrogen sensitivity in the MPOA. *Endocrinology, 144*, 4720–4724.

Champoux, M., Bennett, A., Shannon, C., Higley, J. D., Lesch, K. P., & Suomi, S. J. (2002). Serotonin transporter gene pollymorphism, differential early rearing, and behavior in rhesus monkey neonates. *Molecular Psychiatry, 7*, 1058–1063.

Charalambous, M., da Rocha, S. T., & Ferguson-Smith, A. C. (2007). Genomic imprinting, growth control and the allocation of nutritional resources: Consequences for postnatal life. *Current Opinion in Endocrinology, Diabetes & Obesity, 14*, 3–12.

Chong, S., & Whitelaw, E. (2004). Epigenetic germline inheritance. *Current Opinion in Genetics and Development, 14*, 692–696.

Chrousos, G. P., & Gold, P. W. (1992). The concepts of stress and stress system disorders. *Journal of the American Medical Association, 267*, 1244–1252.

Conger, R., Ge, X., Elder, G., Lorenz, F., & Simons, R. (1994). Economic stress, coercive family process and developmental problems of adolescents. *Child Development, 65*, 541–561.

Conger, R. D., McCarty, J. A., Yang, R. K., Lahey, B. B., & Kropp, J. P. (1984). Perception of child, childrearing values, and emotional distress as mediating links between environmental stressors and observed maternal behavior. *Child Development, 55*, 2234–2247.

Cooney, C. A., Dave, A. A., & Wolff, G. L. (2002). Maternal methyl supplements in mice affect epigenetic variation and DNA methylation of offspring. *Journal of Nutrition, 132*, 2393–2400.

Coplan, J. D., Andrews, M. W., Rosenblum, L. A., & Nemeroff, C. B. (1996). Persistent elevations of cerebrospinal fluid concentrations of corticotropin-releasing factor in adult nonhuman primates exposed to early-life stressors: Implications for the pathophysiology of mood and anxiety disorders. *Proceedings of the National Academy of Sciences, USA, 93*, 1619–1623.

Crosby, S. D., Puetz, J. J., Simburger, K. S., Fahrner, T. J., & Milbrandt, J. (1991). The early response gene NGFI-C encodes a zinc finger transcriptional activator and is a member of the GCGGGGGCG (GSG) element-binding protein family. *Molecular and Cellular Biology, 11*, 3835–3841.

da Rocha, S. T., & Ferguson-Smith, A. C. (2004). Genomic imprinting. *Current Biology, 14*, R646–R649.

Dallman, M. F., Akana, S. F., Strack, A. M., Hanson, E. S., & Sebastian, R. J. (1995). The neural network that regulates energy balance is responsive to glucocorticoids and insulin and also regulates HPA axis responsivity at a site proximal to CRF neurons. *Annals of the New York Academy of Sciences, 771*, 730–742.

Dallman, M. F., Pecoraro, N. C., & la Fleur, S. E. (2005). Chronic stress and comfort foods: Self-medication and abdominal obesity. *Brain Behavior and Immunity, 19*, 275–280.

Davis, M. (2006). Neural systems involved in fear and anxiety measured with fear-potentiated startle. *American Psychologist, 61*, 741–756.

de Kloet, E. R., Karst, H., & Joels, M. (2008). Corticosteroid hormones in the central stress response: Quick-and-slow. *Frontiers in Neuroendocrinology, 29*, 268–272.

Denenberg, V. H. (1964). Critical periods, stimulus input, and emotional reactivity: A theory of infantile stimulation. *Psychological Reviews, 71*, 335–351.

Diamond, M. I., Miner, J. N., Yoshinaga, S. K., & Yamamoto, K. R. (1990). Transcription factor interactions: Selectors of positive or negative regulation from a single DNA element. *Science, 249*, 1266–1270.

Diamond, D. M., & Rose, G. M. (1994). Stress impairs LTP and hippocampal-dependent memory. *Annals of the New York Academy of Sciences, 746*, 411–414.

Dix, T. (1991). The affective organization of parenting adaptive and maladaptive processes. *Psychological Bulletin, 110*, 3–25.

Donaldson, Z. R., & Young, L. J. (2008). Oxytocin, vasopressin, and the neurogenetics of sociality. *Science, 322*, 900–904.

Drevets, W. C., Price, J. L., Simpson, J. R. J., Todd, R. D., Reich, T., Vannier, M., et al. (1997). Subgenual prefrontal cortex abnormalities in mood disorders. *Nature, 386*, 824–827.

Ebstein, R. P. (2003). Relation of shyness in grade school children to the genotype for the long form of the serotonin transporter promoter region polymorphism. *American Journal of Psychiatry, 160*, 671–676.

Ebstein, R. P. (2006). The molecular genetic architecture of human personality: Beyond self-report questionnaires. *Molecular Psychiatry, 11*, 427–445.

Eden, A., Gaudet, F., Waghmare, A., & Jaenisch, R. (2003). Chromosomal instability and tumors promoted by DNA hypomethylation. *Science, 300*, 455.

Eley, T. C., Sugden, K., Corsico, A., Gregory, A. M., Sham, P., McGuffin, P., et al. (2004). Gene–environment interaction analysis of serotonin system markers with adolescent depression. *Molecular Psychiatry, 9*, 908–915.

Etkin, A., Egner, T., Peraza, D. M., Kande, E. R., & Hirsch, J. (2005). Resolving emotional conflict: A role for the rostral anterior cingulate cortex in modulating activity in the amygdala. *Neuron, 51*, 871–882.

Etkin, A., Klemenhagen, K. C., Dudman, J. T., Rogan, M. T., Hen, R., Kandel, E. R., et al. (2004). Individual differences in trait anxiety predict the response of the basolateral amygdala to unconsciously processed fearful faces. *Neuron, 44*, 1043–1055.

Fahrbach, S. E., & Pfaff, D. W. (1986). Effect of preoptic region implants of dilute estradiol on the maternal behavior of ovariectomized, nulliparous rats. *Hormones & Behavior, 20*, 354–363.

Falk, R. (1986). What is a gene? *Studies in the History and Philosophy of Science, 17*, 133–173.

Fanselow, M. S. (2000). Contextual fear, gestalt memories, and the hippocampus. *Behavioural Brain Research, 110*, 73–81.

Farrington, D. A., Gallagher, B., Morley, L., St Ledger, R. J., & West, D. J. (1988). Are there any successful men from criminogenic backgrounds? *Psychiatry, 51*, 116–130.

Feinberg, A. P. (2007). Phenotypic plasticity and the epigenetics of human disease. *Nature, 447*, 433–440.

Felitti, V. J., Anda, R. F., Nordenberg, D., Williamson, D. F., Spitz, A. M., Edwards, V., et al. (1998). Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. *American Journal of Preventive Medicine, 14*, 245–258.

Fisher, P. A., Gunnar, M. R., Chamberlain, P., & Reid, J. B. (2000). Preventive intervention for maltreated preschool children: Impact on children's behavior, neuroendocrine activity and foster parent functioning. *Journal of the American Academy of Child & Adolescent Psychiatry, 39*, 1356–1364.

Fleming, A. S. (1988). Factors influencing maternal responsiveness in humans: Usefulness of an animal model. *Psychoneuroendocrinology, 13*, 189–212.

Fleming, A. S., O'Day, D. H., & Kraemer, G. W. (1999). Neurobiology of mother-infant interactions: Experience and central nervous system plasticity across development and generations. *Neuroscience and Biobehavioral Reviews, 23*, 673–685.

Fox, N. A., Nichols, K. E., Henderson, H. A., Rubin, K., Schmidt, L., Hamer, D., et al. (2005). Evidence for a gene-environment interaction in predicting behavioral inhibition in middle childhood. *Psychological Science, 16*, 921–926.

Francis, D. D., Champagne, F., & Meaney, M. J. (2000). Variations in maternal behaviour are associated with differences in oxytocin receptor levels in the rat. *Journal of Neuroendocrinology, 12*, 1145–1149.

Francis, D. D., Diorio, J., Liu, D., & Meaney, M. J. (1999). Nongenomic transmission across generations in maternal behavior and stress responses in the rat. *Science, 286*, 1155–1158.

Fride, E., Dan, Y., Gavish, M., & Weinstock, M. (1985). Prenatal stress impairs maternal behavior in a conflict situation and reduces hippocampal benzodiazepine receptors. *Life Sciences, 36*, 2103–2109.

Fries, E., Moragues, N., Caldji, C., Hellhammer, H. H., & Meaney, M. J. (2004). Preliminary evidence of altered sensitivity to benzodiazepines as a function of maternal care in the rat. *Annals of the New York Academy of Sciences, 1032*, 320–324.

Gervai, J., Novak, A., Lakatos, K., Toth, I., Danis, I., Ronai, Z., et al. (2007). Infant genotype may moderate sensitivity to maternal affective communications: Attachment disorganization, quality of care, and the DRD4 polymorphism. *Social Neuroscience, 2*, 307–319.

Gillespie, N. A., Whitfield, J. B., Williams, B., Heath, A. C., & Martin, N. G. (2005). The relationship between stressful life events, the serotonin transporter (5-HTTLPR) genotype and major depression. *Psychological Medicine, 35*, 101–111.

Gluckman, P. D., & Hanson, M. A. (2004). Living with the past: Evolution, development, and patterns of disease. *Science, 305*, 1733–1736.

Gluckman, P. D., & Hanson, M. A. (2007). Developmental plasticity and human disease: Research directions. *Journal of Internal Medicine, 261*, 461–471.

Goldstein, L. H., Diener, M. L., & Mangelsdorf, S. C. (1996). Maternal characteristics and social support across the transition to motherhood: Associations with maternal behavior. *Journal of Family Psychology, 10*, 60–71.

Gonzalez, A., Lovic, V., Ward, G. R., Wainwright, P. E., & Fleming, A. S. (2001). Intergenerational effects of complete maternal deprivation and replacement stimulation on maternal behavior and emotionality in female rats. *Developmental Psychobiology, 38*, 11–32.

Gorman, J. M., Kent, J. M., Sullivan, G. M., & Coplan, J. D. (2000). Neuroanatomical hypothesis of panic disorder revised. *American Journal of Psychiatry, 157*, 493–505.

Gottlieb, G. (1991). Experiential canalization of behavioral development: Theory. *Developmental Psychology, 27*, 4–13.

Gottlieb, G. (1997). *Synthesizing nature–nurture*. Mahwah, NJ: Erlbaum.

Gottlieb, G. (1998). Normally occurring environmental and behavioral influences on gene activity: From central dogma to probabilistic epigenesis. *Psychological Reviews, 105*, 792–892.

Grabe, H. J., Lange, M., Wolff, B., Völzke, H., Lucht, M., Freyberger, J. H., et al. (2005). Mental and physical distress is modulated by a polymorphism in the 5-HT transporter gene interacting with social stressors and chronic disease burden. *Molecular Psychiatry, 10*, 220–224.

Grayson, D. R., Jia, X., Chen, Y., Sharma, R. P., Mitchell, C. P., Guidotti, A., et al. (2005). Reelin promoter hypermethylation in schizophrenia. *Proceedings of the National Academy of Sciences, USA, 102*, 9341–9346.

Greenberg, B. D., Tolliver, T. J., Huang, S. J., Li, Q., Bengel, D., & Murphy, D. L. (1999). Genetic variation in the serotonin transporter promoter region affects serotonin uptake in human blood platelets. *American Journal of Medical Genetics, 88*, 83–87.

Griffiths, P. E., & Stotz, K. (2007). Gene. In M. Ruse & D. Hull (Eds.), *Cambridge companion to philosophy of biology* (pp. 85–102). Cambridge, UK: Cambridge University Press.

Griffiths, P. E., & Tabery, J. (2008). Behavioral genetics and development: Historical and conceptual causes of controversy. *New Ideas in Psychology, 26*, 332–352.

Grindstaff, J. L., Brodie, E. D., & Ketterson, E. D. (2003). Immune function across generations: Integrating mechanism and evolutionary process in maternal antibody transmission. *Proceedings of the Royal Society B: Biological Sciences, 270*, 2309–2319.

Grolnick, W. S., Gurland, S. T., DeCourcey, W., & Jacob, K. (2002). Antecedents and consequences of mothers' autonomy support: An experimental investigation. *Developmental Psychology, 38*, 143–155.

Gross, C., & Hen, R. (2004). The developmental origins of anxiety. *Nature Reviews Neuroscience, 5*, 545–552.

Gross, C., Zhuang, X., Stark, K., Ramboz, S., Oosting, R., Kirby, L., et al. (2002). Serotonin1A receptor acts during development to establish normal anxiety-like behaviour in the adult. *Nature, 416*, 396–400.

Grunstein, M. (1997). Histone acetylation in chromatin structure and trasncription. *Nature, 389*, 349–352.

Haapasap, J., & Tremblay, R. E. (1994). Physically aggressive boys from ages 6 to 12: Family background. Parenting behavior, and prediction of delinquency. *Journal of Consulting & Clinical Psychology, 62*, 1044–1052.

Hake, S. B., & Allis, C. D. (2006). Histone H3 variants and their potential role in indexing mammalian genomes: The "H3 barcode hypothesis." *Proceedings of the National Academy of Sciences, USA, 103*, 6428–6435.

Hammock, E. A. D., & Young, L. J. (2005). Microsatellite instability generates diversity in brain and sociobehavioral traits. *Science, 308*, 1630–1634.

Hane, A. A., & Fox, N. A. (2006). Ordinary variations in maternal caregiving influence human infants' stress reactivity. *Psychological Sciences, 17*, 550–556.

Hariri, A. R., Drabant, E. M., Munoz, K. E., Kolachana, B. S., Mattay, V. S., Egan, M. F., et al. (2005). A susceptibility gene for affective disorders and the response of the human amygdala. *Archives of General Psychiatry, 62*, 146–152.

Hariri, A. R., Drabant, E. M., & Weinberger, D. R. (2006). Imaging genetics: Perspectives from studies of genetically driven variation in serotonin function and corticolimbic affective processing. *Biological Psychiatry, 59*, 888–897.

Hariri, A. R., Mattay, V. S., Tessitore, A., Kolachana, B., Fera, F., Goldman, D., et al. (2002). Serotonin transporter genetic variation and the response of the human amygdala. *Science, 297*, 400–403.

Hasselquist, D., & Nilsson, J.-A. (2009). Maternal transfer of antibodies in vertebrates: Trans-generational effects on offspring immunity. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*, 51–60.

Heim, C., & Nemeroff, C. B. (2001). The role of childhood trauma in the neurobiology of mood and anxiety disorders: Preclinical and clinical studies. *Biological Psychiatry, 49*, 1023–1039.

Heim, C., Newport, D. J., Heit, S., Graham, Y. P., Wilcox, M., Bonsall, R., et al. (2000). Pituitary–adrenal and autonomic responses to stress in women after sexual and physical abuse in childhood. *Journal of the American Medical Association, 284*, 592–597.

Heinz, A., Braus, D. F., Smolka, M. N., Wrase, J., Puls, I., Hermann, D., et al. (2005). Amygdala-prefrontal coupling depends on a genetic variation of the serotonin transporter. *Nature Neuroscience, 8*, 20–21.

Heitzer, M. D., Wolf, I. M., Sanchez, E. R., Witchel, S. F., & DeFranco, D. B. (2007). Glucocorticoid receptor physiology. *Reviews in Endocrine & Metabolic Disorders, 8*, 321–330.

Hellman, A., & Chess, A. (2007). Gene body-specific methylation on the active X chromosome. *Science, 315*, 1141–1143.

Hellstrom, I. C., Zhang, T. Y., Diorio, J., & Meaney, M. J. (2009). *Maternal regulation of glucocorticoid receptor expression through serotonin-induced nerve growth factor-inducible factor A signaling*. Unpublished manuscript.

Hettema, J. M., Neale, M. C., Myers, J. M., Prescott, C. A., & Kendler, K. S. (2006). A population-based twin study of the relationship between neuroticism and internalizing disorders. *American Journal of Psychiatry, 163*, 857–864.

Hinde, R. A. (1986). Some implications of evolutionary theory and comparative data for the study of human prosocial and aggressive behaviour. In D. Olweus, J. Block, & M. Radke-Yarrow (Eds.), *Development of antisocial and prosocial behaviour* (pp. 13–32). Orlando, FL: Academic Press.

Hofer, M. A. (2005). The psychobiology of early attachment. *Clinical Neuroscience Research, 4*, 291–300.

Holliday, R. (1989). DNA methylation and epigenetic mechanisms. *Cell Biophysics, 15*, 15–20.

Holsboer, F. (2000). The corticosteroid receptor hypothesis of depression. *Neuropsychopharmacology, 23*, 477–501.

Hu, X., Oroszi, G., Chun, J., Smith, T. L., Goldman, D., & Schuckit, M. A. (2005). An expanded evaluation of the relationship of four alleles to the level of response to alcohol and the alcoholism risk. *Alcohol Clinical Experimental Research, 29*, 8–16.

Jablonka, E., & Lamb, M. J. (2005). *Evolution in four dimensions: Genetic, epigenetic, behavioral, and symbolic variation in the history of life*. Cambridge, MA: MIT Press.

Jenuwein, T., & Allis, C. D. (2001). Translating the histone code. *Science, 293*, 1074–1080.

Jiang, Y., Langley, B., Lubin, F. D., Renthal, W., Wood, M. A., Yasui, D. H., et al. (2008). Epigenetics in the nervous system. *Journal of Neuroscience, 28*, 11753–11759.

Jirtle, R. L., & Skinner, M. K. (2007). Environmental epigenomics and disease susceptibility. *Nature Reviews Genetics, 8*, 253–262.

Joels, M., Karst, H., Krugers, H. J., & &. Lucassen, P. J. (2007). Chronic stress: Implications for neuronal morphology, function and neurogenesis. *Frontiers in Neuroendocrinology, 28*, 72–96.

Johannsen, W. (1911). The genotype conception of heredity. *The American Naturalist, 45*, 129–159.

Jokela, M., Keltikangas-Jarvinen, L., Kivimaki, M., Puttonen, S., Elovainio, M., Rontu, R., et al. (2007). Serotonin receptor 2A gene and the influence of childhood maternal nurturance on adulthood depressive symptoms. *Archives of General Psychiatry, 64*, 356–360.

Jutapakdeegul, N., Casalotti, S. O., Govitrapong, P., & Kotchabhakdi, N. (2003). Postnatal touch stimulation acutely alters corticosterone levels and glucocorticoid receptor gene expression in the neonatal rat. *Developmental Neuroscience, 25*, 26–33.

Kaufman, J., Yang, B. Z., Douglas-Palumberi, H., Houshyar, S., Lipschitz, D., Krystal, J. H., et al. (2004). Social supports and serotonin transporter gene moderate depression in maltreated children. *Proceedings of the National Academy of Sciences, USA, 101*, 17316–17321.

Kendler, K. S., Kuhn, J., & Prescott, C. A. (2004). The interrelationship of neuroticism, sex, and stressful life events in the prediction of episodes of major depression. *American Journal of Psychiatry, 161*, 631–636.

Kendler, K. S., Kuhn, J. W., Vittum, J., Prescott, C. A., & Riley, B. (2005). The interaction of stressful life events and a serotonin transporter polymorphism in the prediction of episodes of major depression: A replication. *Archives of General Psychiatry, 62*, 529–535.

Kim, J. J., & Yoon, K. S. (1998). Stress: Metaplastic effects in the hippocampus. *Trends in Neurosciences, 21*, 505–509.

Kinsley, C. H., Mann, P. E., & Bridges, R. S. (1998). Prenatal stress alters morphine- and stress induced analgesia in male and female rats. *Pharmacology, Biochemistry and Behavior, 50*, 413–419.

Kirkwood, A., Dudek, S. M., Gold, J. T., Aizenman, C. D., & Bear, M. F. (1993). Common forms of synaptic plasticity in the hippocampus and neocortex in vitro. *Science, 260*, 1518–1521.

Klose, R. J., & Bird, A. P. (2006). Genomic DNA methylation: The mark and its mediators. *Trends in Biochemical Sciences, 31*, 89–97.

Koob, G. F., Heinrichs, S. C., Menzaghi, F., Pich, E. M., & Britton, K. T. (1994). Corticotropin-releasing factor, stress and behavior. *Seminars in Neuroscience, 6*, 221–229.

Kumar, R., & Thompson, E. B. (2005). Gene regulation by the glucocorticoid receptor: Structure–function relationship. *Journal of Steroid Biochemistry & Molecular Biology, 94*, 383–394.

LaBar, K. S., Gatenby, J. C., LeDoux, J. E., et al. (1998). Human amygdala activation during conditioned fear acquisition and extinction—A mixed-trial fMRI study. *Neuron, 205*, 937–945.

LaBar, K. S., LeDoux, J. E., Spencer, D. D., et al. (1995). Impaired fear conditioning following unilateral temporal lobectomy in humans. *Journal of Neuroscience, 15*, 6846–6855.

Laird, P. W. (2005). Cancer epigenetics. *Human Molecular Genetics, 14*(Spec. No. 1), R65–R76.

Lakatos, K., Nemoda, Z., Birkas, E., Ronai, Z., Kovacs, E., Ney, K., et al. (2003). Association of D4 dopamine receptor gene and serotonin transporter promoter polymorphisms with infants' response to novelty. *Molecular Psychiatry, 8*, 90–97.

Lakatos, K., Nemoda, Z., Toth, I., Ronai, Z., Ney, K., Sasvari- Szekely, M., et al. (2002). Further evidence for the role of the dopamine D4 receptor (DRD4) gene in attachment disorganization: Interaction of the exon III 48-bp repeat and the -521 C/T promoter polymorphisms. *Molecular Psychiatry, 7*, 27–31.

LeDoux, J. E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience, 23*, 155–184.

Lee, M. H., & Williams, D. I. (1977). A longitudinal study of mother-young interaction in the rat: The effects of infantile stimulation, diurnal rhythms, and pup maturation. *Behaviour, 63*, 241–261.

Lesch, K., Bengel, D., Heils, A., Sabol, S., Greenberg, B., Petri, S., et al. (1996). Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science, 274*, 1527–1531.

Lesch, K. P., Meyer, J., Glatz, K., Flü gge, G., Hinney, A., Hebebrand, J., et al. (1997). The 5-HT transporter gene-linked polymorphic region (5-HTTLPR) in evolutionary perspective: Alternative biallelic variation in rhesus monkeys. *Journal of Neural Transmission, 104*, 1259–1266.

Leserman, J., Drosman, D. A., Li, Z., Toomey, T. C., Nachman, G., & Glogau, L. (1996). Sexual and physical abuse history in gasteroenterology practice: How types of abuse impact health status. *Psychosomatic Medicine, 58*, 4–15.

Levine, S. (1970). The pituitary–adrenal system and the developing brain. *Progress in Brain Research, 32*, 79–85.

Levine, S. (1994). The ontogeny of the hypothalamic–pituitary–adrenal axis. The influence of maternal factors. *Annals of the New York Academy of Sciences, 746*, 275–288.

Lewontin, R. C. (1974). The analysis of variance and the analysis of causes. *American Journal of Human Genetics, 26*, 400–411.

Lissau, J., & Sorensen, T. I. A. (1994). Parental neglect during childhood and increased risk of obesity in young adulthood. *Lancet, 343*, 324–327.

Liu, D., Diorio, J., Day, J. C., Francis, D. D., Mar, A., & Meaney, M. J. (2000). Maternal care, hippocampal synaptogenesis and cognitive development in the rat. *Nature Neuroscience, 3*, 799–806.

Liu, D., Diorio, J., Tannenbaum, B., Caldji, C., Francis, D. D., Freedman, A., et al. (1997). Maternal care, hippocampal glucocorticoid receptors and HPA responses to stress. *Science, 277*, 1659–1662.

Lovic, V., Gonzalez, A., & Fleming, A. S. (2001). Maternally separated rats show deficits in maternal care in adulthood. *Developmental Psychobiology, 39*, 19–33.

Lu, N. Z., & Cidlowski, J. A. (2006). Glucocorticoid receptor isoforms generate transcription specificity. *Trends in Cell Biology, 16*, 301–307.

Lubin, F. D., Roth, T. L., & Sweatt, J. D. (2008). Epigenetic regulation of *bdnf* gene transcription in the consolidation of fear memory. *Journal of Neuroscience, 28*, 10576–10586.

Lucki, I. (1998). The spectrum of behaviors influenced by serotonin. *Biological Psychiatry, 44*, 151–162.

Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome coreparticle at 2.8 A resolution. *Nature, 389*, 251–260.

Maestripieri, D. (2005). Early experience affects the intergenerational transmission of infant abuse in rhesus monkeys. *Proceedings of the National Academy of Sciences, USA, 102*, 9726–9729.

Maren, S., & Quirk, G. J. (2004). Neuronal signalling of fear memory. *Nature Reviews Neuroscience, 5*, 844–852.

Martinowich, K., Hattori, D., Wu, H., Fouse, S., He, F., Hu, Y., et al. (2003). DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. *Science, 302*, 890–893.

McCauley, J., Kern, D. E., Kolodner, K., Dill, L., Schroeder, A. F., DeChant, H. K., et al. (1997). Clinical characteristics of women with a history of childhood abuse: Unhealed wounds. *Journal of the American Medical Association, 277*, 1362–1368.

McCormick, J. A., Lyons, V., Jacobson, M. D., Noble, J., Diorio, J., Nyirenda, M., et al. (2000). 5'Heterogeneity of glucocorticoid receptor messenger ribonucleic acid is tissue-specific: Differential regulation of variant transcripts by early life events. *Molecular Endocrinology, 14*, 506–517.

McEwen, B. S. (2003). Early life influences on life-long patterns of behavior and health. *Mental Retardation & Developmental Disabilities Research Reviews, 9*, 149–154.

McEwen, B. S. (2007). Physiology and neurobiology of stress and adaptation: Central role of the brain. *Physiological Reviews, 87*, 873–904.

McGirr, A., Renaud, J., Seguin, M., Alda, M., & Turecki, G. (2008). Course of major depressive disorder and suicide outcome: A psychological autopsy study. *Journal of Clinical Psychiatry, 69*, 966–970.

McGowan, P. O., Sasaki, A., Dymov, S., LaBoonté, B., Turecki, G., Szyf, M., et al. (2009). Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. *Nature Neuroscience, 12*, 342–348.

McLloyd, V. C. (1998). Socioeconomic disadvantage and child development. *American Psychologist, 53*, 185–204.

McNally, J. G., Müller, W. G., Walker, D., Wolford, R., & Hager, G. L. (2000). The glucocorticoid receptor: Rapid exchange with regulatory sites in living cells. *Science, 287*, 1262–1265.

Meaney, M. J. (2001). The development of individual differences in behavioral and endocrine responses to stress. *Annual Review of Neuroscience, 24*, 1161–1192.

Meaney, M. J. (2007). Social influences on sexual differentiation in mammals. *Advances in Genetics, 59,* 173–215.

Meaney, M. J., Diorio, J., Donaldson, L., Yau, J., Chapman, K., & Seckl, J. R. (2000). Handling alters the expression of messenger RNAs for AP-2, NGFI-A and NGFI-B in the hippocampus of neonatal rats. *Journal of Neuroscience, 20,* 3936–3945.

Meaney, M. J., LeDoux, J. E., & Leibowitz, M. L. (2008). Neurobiology of anxiety disorders. In A. Tasman, J. Kay, J. A. Lieberman, M. B. First, & M. Maj (Eds.). *Psychiatry, 1,* 317–338.

Meaney, M. J., & Szyf, M. (2005). Maternal effects as a model for environmentally-dependent chromatin plasticity. *Trends in Neuroscience, 28,* 456–463.

Meaney, M. J., Szyf, M., & Seckl, J. R. (2007). Epigenetic mechanisms of perinatal programming of hypothalamic–pituitary–adrenal function and health. *Trends in Molecular Medicine, 13,* 269–277.

Melo, A. I., Lovic, V., Gonzalez, A., Madden, M., Sinopoli, K., & Fleming, A. S. (2006). Maternal and littermate deprivation disrupts maternal behavior and social-learning of food preference in adulthood: Tactile stimulation, nest odor, and social rearing prevent these effects. *Developmental Psychobiology, 48,* 209–219.

Menard, J., Champagne, D., & Meaney, M. J. (2004). Maternal care alters behavioral and neural activity patterns in the defensive burying paradigm. *Neuroscience, 129,* 297–308.

Miesfield, R., Okret, S., Wikstrom, A.-C., Wrange, O., Gustafsson, A. J., & Yamamoto, K. R. (1984). Characterization of a steroid hormone receptor gene and mRNA in wild-type and mutant cells. *Nature, 312,* 779–781.

Miller, C. A., & Sweatt, J. D. (2007). Covalent modification of DNA regulates memory formation. *Neuron, 53,* 857–869.

Mitchell, J. B., Betito, K., Rowe, W., Boksa, P., & Meaney, M. J. (1992). Serotonergic regulation of type II corticosteroid receptor binding in cultured hippocampal cells: The role of serotonin-induced increases in cAMP levels. *Neuroscience, 48,* 631–639.

Mitchell, J. B., Rowe, W., Boksa, P., & Meaney, M. J. (1990). Serotonin regulates type II, corticosteroid receptor binding in cultured hippocampal cells. *Journal of Neuroscience, 10,* 1745–1752.

Mohandas, T., Sparkes, R. S., & Shapiro, L. J. (1981). Reactivation of an inactive human X chromosome: Evidence for X inactivation by DNA methylation. *Science, 211,* 393–396.

Moore, C. L., & Power, K. L. (1986). Prenatal stress affects mother-infant interaction in Norway rats. *Developmental Psychobiology, 19,* 235–245.

Morris, J. S., Ohman, A., & Dolan, R. J. (1998). Conscious and unconscious emotional learning in the human amygdala. *Nature, 393,* 467–470.

Moss, L. (2005). *What genes can't do.* Cambridge, MA: MIT Press.

Mousseau, T. A., & Fox, C. W. (1998). The adaptive significance of maternal effects. *Trends in Ecology and Evolution, 13,* 403–407.

Munafo, M. R., Clark, T., & Flint, J. (2005). Does measurement instrument moderate the association between the serotonin transporter gene and anxiety related personality traits? A meta-analysis. *Molecular Psychiatry, 10,* 415–419.

Munck, A. (2005). Glucocorticoid receptors and physiology: A personal history. *Steroids, 70,* 335–344.

Munck, A., Guyre, P. M., & Holbrook, N. J. (1984). Physiological functions of glucocorticoids in stress and their relation to pharmacological actions. *Endocrine Reviews, 5,* 25–44.

Murayama, A., Sakura, K., Nakama, M., Yasuzawa-Tanaka, K., Fujita, E., Tateishi, Y., et al. (2006). A specific CpG site demethylation in the human interleukin 2 gene promoter is an epigenetic memory. *EMBO Journal, 25,* 1081–1092.

Neigh, G. N., & Nemeroff, C. B. (2006). Reduced glucocorticoid receptors: Consequence or cause of depression? *Trends in Endocrinology & Metabolism, 17,* 124–125.

Nissen, R. M., & Yamamoto, K. R. (2000). The glucocorticoid receptor inhibits NFkappaB by interfering with serine-2 phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes & Development, 14,* 2314–2329.

Nobukuni, Y., Smith, C. L., Hager, G. L., & Detera-Wadleigh, S. D. (1995). Characterization of the human glucocorticoid receptorpromoter. *Biochemistry, 34,* 8207–8214.

Novakov, M., & Fleming, A. S. (2005). The effects of early rearing environment on the hormonal induction of maternal behavior in virgin rats. *Hormones & Behavior, 48,* 528–536.

Numan, M. J., & Insel, T. R. (2003). *The neurobiology of parental behavior* (3rd ed.). New York: Springer.

Offord, D. R., Boyle, M. H., Racine, Y. A., Fleming, J. E., Cadman, D. T., Blum, H. M., et al. (1992). Outcome, prognosis, and risk in a longitudinal follow-up study. *Journal of the American Academy of Child and Adolescent Psychiatry, 31,* 916–923.

Olds, D., Henderson, C. R., Cole, R., Eckenrode, J., Kitzman, H., Luckey, D., et al. (1998). Long-term effects of nurse home visitation on children's criminal and antisocial behavior: 15-year follow-up of a randomized controlled trial. *Journal of the American Medical Association, 280,* 1238–1244.

Pedersen, C. A. (1997). Oxytocin control of maternal behavior. Regulation by sex steriods and offspring stimuli. *Annals of the New York Academy of Sciences, 614,* 126–145.

Pérez-Edgar, K., & Fox, N. A. (2005). Temperament and anxiety disorders. *Child and Adolescent Psychiatric Clinics of North America, 14,* 681–706.

Pezawas, L., Meyer-Lindenberg, A., Drabant, E. M., Verchinski, B. A., Munoz, K. E., Kolachana, B. S., et al. (2005). 5-HTTLPR polymorphism impacts human cingulate-amygdala interactions: A genetic susceptibility mechanism for depression. *Nature Neuroscience, 8*, 828–834.

Plomin, R., DeFries, J. C., & McClearn, G. E. (1990). *Behavioral genetics. A primer*. New York: Freeman.

Plomin, R., & Rutter, M. (1998). Child development, molecular genetics, and what to do with genes once they are found. *Child Development, 69*, 1223–1242.

Plotsky, P. M., Cunningham, E. T., Jr., & Widmaier, E. P. (1989). Catecholaminergic modulation of corticotropin-releasing factor and adrenocorticotropin secretion. *Endocrine Reviews, 10*, 437–458.

Propper, C., Willoughby, M., Halpern, C. T., Carbone, M. A., & Cox, M. (2007). Parenting quality, DRD4, and the prediction of externalizing and internalizing behaviors in early childhood. *Developmental Psychobiology, 49*, 619–632.

Pruessner, J. L., Champagne, F. A., Meaney, M. J., & Dagher, A. (2004). Parental care and neuroendocrine and dopamine responses to stress in humans: A PET imaging study. *Journal of Neuroscience, 24*, 2825–2831.

Ptak, C., & Petronis, A. (2008). Epigenetics and complex disease: From etiology to new therapeutics. *Annual Review of Pharmacology & Toxicology, 48*, 257–276.

Raison, C. L., & Miller, A. H. (2003). When not enough is too much: The role of insufficient glucocorticoid signaling in the pathophysiology of stress-related disorders. *American Journal of Psychiatry, 160*, 1554–1565.

Rassoulzadegan, M., Grandjean, V., Gounon, P., Vincent, S., Gillot, I., & Cuzin, F. (2006). RNA-mediated non-mendeliaqn inheritance of an epigenetic change in the mouse. *Nature, 441*, 469–474.

Razin, A. (1998). CpG methylation, chromatin structure and gene silencing-a three-way connection. *EMBO Journal, 17*, 4905–4908.

Razin, A., & Cedar, H. (1993). DNA methylation: Molecular biology and biological significance. In J. P. Jost & H. P. Saluz (Eds.), *DNA methylation: Molecular biology and biological significance* (pp. 343–359). Basel, Switzerland: Birkhäuser Verlag.

Razin, A., & Riggs, A. D. (1980). DNA methylation and gene function. *Science, 210*, 604–610.

Reik, W. J. (2001). Genomic imprinting: Parental influence on the genome. *Nature, 293*, 1089–1096.

Renthal, W., Maze, I., Krishnan, V., Covington, H. E., III, Xiao, G., Kumar, A., et al. (2007). Histone deacetylase 5 epigenetically controls behavioral adaptations to chronic emotional stimuli. *Neuron, 56*, 517–529.

Renthal, W., & Nestler, E. J. (2008). Epigenetic mechanisms in drug addiction. *Trends in Molecular Medicine, 14*, 341–350.

Repetti, R. L., Taylor, S. E., & Seeman, T. E. (2002). Risky families: Family social environments and the mental and physical health of offspring. *Psychological Bulletin, 128*, 330–366.

Riggs, A. D., & Pfeifer, G. P. (1992). X-chromosome inactivation and cell memory. *Trends in Genetics, 8*, 169–174.

Robinson, G. E., Fernald, R. D., & Clayton, D. F. (2008). Genes and social behavior. *Science, 322*, 896–900.

Rosen, J. B., & Schulkin, J. (1998). From normal fear to pathological anxiety. *Psychological Review, 105*, 325–350.

Rossiter, M. C. (1998). The role of environmental variation in parental effects expression. In T. A. Mousseau & C. W. Fox (Eds.), *Maternal effects as adaptations* (pp. 112–136). Oxford, UK: Oxford University Press.

Roth, T. L., Lubin, F. D., Funk, A. J., & Sweatt, J. D. (2009). Lasting epigenetic influence of early-life adversity on the BDNF gene. *Biological Psychiatry, 65*, 760–769.

Russak, L. G., & Schwartz, G. E. (1997). Feelings of parental care predict health status in midlife: A 35 year follow-up of the Harvard Mastery of Stress Study. *Journal of Behavioral Medicine, 20*, 1–11.

Rutter, M. (2007). Gene–environment interdependence. *Developmental Science, 10*, 12–18.

Sapolsky, R. M., & Meaney, M. J. (1986). The maturation of the adrenocortical stress response in the rat. *Brain Research Reviews, 11*, 65–76.

Sapolsky, R. M., Romero, L. M., & Munck, A. U. (2000). How do glucocorticoids influence stress responses? Integrating permissive, suppressive, stimulatory, and preparative actions. *Endocrine Reviews, 21*, 55–89.

Schanberg, S. M., Evoniuk, G., & Kuhn, C. M. (1984). Tactile and nutritional aspects of maternal care: Specific regulators of neuroendocrine function and cellular development. *Proceedings of the Society for Experimental Biology & Medicine, 175*, 135–146.

Schatz, C. J. (1990). Impulse activity and the patterning of connections during CNS development. *Neuron, 5*, 745–756.

Schatzberg, A. F., Rothschild, A. J., Langlais, P. J., Bird, E. D., & Cole, J. O. (1985). A corticosteroid/dopamine hypothesis for psychotic depression and related states. *Journal of Psychiatric Research, 19*, 57–64.

Schinka, J. A., Busch, R. M., & Robichaux-Keene, N. (2004). A meta-analysis of the association between the serotonin transporter gene polymorphism (5-HTTLPR) and trait anxiety. *Molecular Psychiatry, 9*, 197–202.

Schneirla, T. C. (1966). Behavioral development and comparative psychology. *Quarterly Review of Biology, 41*, 283–302.

Seckl, J. R. (2001). Glucocorticoid programming of the fetus adult phenotypes and molecular mechanisms. *Molecular and Cellular Endocrinology, 185*, 61–71.

Seckl, J. R., & Holmes, M. C. (2007). Mechanisms of disease: Glucocorticoids, their placental metabolism and fetal "programming" of adult pathophysiology. *Nature Clinical Practice Endocrinology & Metabolism, 3*, 479–488.

Seckl, J. R., & Meaney, M. J. (1994). Early life events and later development of ischaemic heart disease. *Lancet, 342*, 1236.

Sen, S., Burmeister, M., & Ghosh, D. (2004). Meta-analysis of the association between a serotonin transporter promoter polymorphism (5-HTTLPR) and anxiety-related personality traits. *American Journal of Medical Genetics, 127B*, 85–89.

Shahrohk, D., Cameron, N. M., Zhang, T.-Y., Gratton, A., & Meaney, M. J. (in press). Oxytocin–dopamine interactions in the ventral tegmental area mediate variations in maternal behaviour in the rat.

Shanks, N., Larocque, S., & Meaney, M. J. (1995). Neonatal endotoxin exposure alters the development of the hypothalamic–pituitary–adrenal axis: Early illness and later responsivity to stress. *Journal of Neuroscience, 15,* 376–384.

Shanks, N., Windle, R. J., Perks, P. A., Harbuz, M. S., Jessop, D. S., Ingram, C. D., et al. (2000). Early-life exposure to endotoxin alters hypothalamic-pituitary-adrenal function and predisposition to inflammation. *Proceedings of the National Academy of Sciences, USA, 97,* 5645–5650.

Shin, L. M., Orr, S. P., Carson, M. A., Rauch, S. L., Macklin, M. L., Lasko, N. B., et al. (2004). Regional cerebral blood flow in the amygdala and medial prefrontal cortex during traumatic imagery in male and female Vietnam veterans with PTSD. *Archives of General Psychiatry, 61,* 168–176.

Shin, L. M., Wright, C. I., Cannistraro, P. A., Wedig, M. M., McMullin, K., Martis, B., et al. (2005). A functional magnetic resonance imaging study of amygdala and medial prefrontal cortex responses to overtly presented fearful faces in posttraumatic stress disorder. *Archives of General Psychiatry, 62,* 273–281.

Shors, T. J., Foy, M. R., Levine, S., & Thompson, R. F. (1990). Unpredictable and uncontrollable stress impairs neuronal plasticity in the rat hippocampus. *Brain Research Bulletin, 24,* 663–667.

Shorter, J., & Lindquist, S. (2005). Prions as adaptive conduits of memory and inheritance. *Nature Reviews Genetics, 6,* 435–450.

Smith, J. W., Seckl, J. R., Evans, A. T., Costall, B., & Smythe, J. W. (2004). Gestational stress induces post-partum depression-like behaviour and alters maternal care in rats. *Psychoneuroendocrinology, 29,* 227–244.

Sokolowski, M. B., & Wahlsten, D. (2001). Gene–environment interaction and complex behavior. In S. O. Moldin (Ed.), *Methods in genomic neuroscience* (pp. 3–27). Boca Raton, FL: CRC Press.

Stutzmann, G. E., & LeDoux, J. E. (1999). GABAergic antagonists block the inhibitory effects of serotonin in the lateral amygdala: A mechanism for modulation of sensory inputs related to fear conditioning. *Journal of Neuroscience, 19,* RC8.

Suomi, S. J. (2006). Risk, resilience, and gene × environment interactions in rhesus monkeys. *Annals of the New York Academy of Sciences, 1094,* 83–104.

Surtees, P. G., Wainwright, N. W. J., Saffron, A. G., Lubena, R., Day, N. E., & Flint, J. (2006). Social adversity, the serotonin transporter (5-HTTLPR) polymorphism and major depressive disorder. *Biological Psychiatry, 59,* 224–229.

Sweatt, J. D. (2009). Experience-dependent epigenetic modifications in the central nervous system. *Biological Psychiatry, 65,* 191–197.

Szyf, M., Weaver, I. C., Champagne, F. A., Diorio, J., & Meaney, M. J. (2005). Maternal programming of steroid receptor expression and phenotype through DNA methylation in the rat. *Frontiers in Neuroendocrinology, 26,* 139–162.

Tang, A. C., Akers, K. G., Reeb, B. C., Romeo, R. D., & McEwen, B. S. (2006). Programming social, cognitive, and neuroendocrine development by early exposure to novelty. *Proceedings of the National Academy of Sciences, USA, 103,* 15716–15721.

Toki, S., Morinobu, S., Imanaka, A., Yamamoto, S., Yamawaki, S., & Honma, K. (2007). Importance of early lighting conditions in maternal care by dam as well as anxiety and memory later in life of offspring. *European Journal of Neuroscience, 25,* 815–829.

Tsankova, N., Renthal, W., Kumar, A., & Nestler, E. J. (2007). Epigenetic regulation in psychiatric disorders. *Nature Reviews Neuroscience, 8,* 355–367.

Turner, B. (2001). *Chromatin structure and the regulation of gene expression.* Oxford, UK: Blackwell.

Turner, J. D., & Muller, C. P. (2005). Structure of the glucocorticoid receptor (NR3C1) gene 5′untranslated region: Identification and tissue distribution of multiple new human exon 1. *Journal of Molecular Endocrinology, 35,* 283–292.

Uno, H., Tarara, R., Else, G., Suleman, M. A., & Sapolsky, R. M. (1989). Hippocampal damage associated with prenatal glucocorticoid exposure. *Developmental Brain Research, 9,* 1705–1711.

Valentino, R. J., Curtis, A. L., Page, M. E., Pavcovich, L. A., & Florin-Lechner, S. M. (1998). Activation of the locus cereulus brain noradrenergic system during stress: Circuitry, consequences, and regulation. *Advances in Pharmacology, 42,* 781–784.

Van den Boom, D. (1994). The influence of temperament and mothering on attachment and exploration: An experimental manipulation of sensitive responsiveness among lower-class mothers and irritable infants. *Child Development, 65,* 1457–1477.

Vardhman, K., Rakyan, V. K., Suyinn Chong, S., Marnie, E., Champ, M. E., Cuthbert, P. C., et al. (2003). Transgenerational inheritance of epigenetic states at the murine AxinFu allele occurs after maternal and paternal transmission. *Proceedings of the National Academy of Sciences, USA, 100,* 2538–2543.

Vaughn, B., Egeland, B., Sroufe, L. A., & Waters, E. (1979). Individual differences in infant-mother attachment at twelve and eighteen months: Stability and change in families under stress. *Child Development, 50,* 971–975.

Wahlsten, D. (1990). Insensitivity of the analysis of variance to heredity–environment interaction. *Behavioral and Brain Sciences, 13,* 109–161.

Wahlsten, D. (2003). Airbrushing heritability. *Genes, Brain, & Behavior, 2,* 327–329.

Walker, B. R. (2007). Glucocorticoids and cardiovascular disease. *European Journal of Endocrinology, 157,* 545–559.

Waterland, R. A., & Jirtle, R. L. (2003). Transposable elements: Targets for early nutritional effects on epigenetic gene regulation. *Molecular Cell Biology, 23,* 5293–5300.

Waterland, R. A., Lin, J. R., Smith, C. A., & Jirtle, R. L. (2006). Post-weaning diet affects genomic imprinting at the insulin-like growth factor 2 (Igf2) locus. *Human Molecular Genetics, 15,* 705–716.

Weaver, I. C. G., Cervoni, N., D'Alessio, A. C., Champagne, F. A., Seckl, J. R., Szyf, M., et al. (2004). Epigenetic programming through maternal behavior. *Nature (Neuroscience), 7,* 847–854.

Weaver, I. C. G., Champagne, F. A., Brown, S. E., Dymov, S., Sharma, S., Meaney, M. J., et al. (2005). Reversal of maternal programming of stress responses in adult offspring through methyl supplementation: Altering epigenetic marking later in life. *Journal of Neuroscience, 25,* 11045–11054.

Weaver, I. C. G., DiAlessio, A. C., Brown, S. E., Hellstrom, I. C., Dymov, S., Diorio, J., et al. (2007). The transcription factor NGFI-A mediates epigenetic programming: Altering epigenetic marking through immediate early genes. *Journal of Neuroscience, 27,* 1756–1768.

Weber, M., Hellmann, I., Stadler, M. B., Ramos, L., Paabo, S., Rebhan, M., et al. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature Genetics, 39,* 457–466.

West, M. J., & King, A. P. (1987). Settling nature and nurture into an ontogenetic niche. *Developmental Psychobiology, 20,* 549–562.

Whitelaw, N. C., & Whitelaw, E. (2006). How lifetimes shape epigeneotpye within and across generations. *Human Molecular Genetics, 15,* R131–R137.

Yehuda, R., & Bierer, L. M. (2008). Transgenerational transmission of cortisol and PTSD risk. *Progress in Brain Research, 167,* 121–135.

Zhang, T.-Y., Bagot, R., Parent, C., Nesbitt, C., Bredy, T. W., Caldji, C., et al. (2006). Maternal programming of defensive responses through sustained effects on gene expression. *Biological Psychiatry, 73,* 72–89.

Zhang, T., Haws, P., & Wu, Q. (2004). Multiple variable first exons: A mechanism for cell- and tissue-specific gene regulation. *Genome Research, 14,* 79–89.

Zhang, T. Y., Hellstrom, I. C., Wei, X. L., & Meaney, M. J. (2009). *Maternal regulation of glutamic acid decarboxylase 1 gene expression through DNA methylation.* Manuscript submitted for publication.

Child Development, January/February 2010, Volume 81, Number 1, Pages 80–100

A Quasi-Experimental Study of Maternal Smoking During Pregnancy and Offspring Academic Achievement

Brian M. D'Onofrio and
Amber L. Singh
Indiana University

Anastasia Iliadou, Mats Lambe,
and Christina M. Hultman
Karolinska Institutet

Jenae M. Neiderhiser
The Pennsylvania State University

Niklas Långström and
Paul Lichtenstein
Karolinska Institutet

The current study, based on all births in Sweden from 1983 to 1991 ($N = 654{,}707$), explored the processes underlying the association between smoking during pregnancy (SDP) and offspring school grades and mathematic proficiency at age 15. The analyses compared relatives who varied in their exposure to SDP and who varied in their genetic relatedness. Although SDP was statistically associated with academic achievement (AA) when comparing unrelated individuals, the results suggest that SDP does not cause poorer academic performance, as full siblings differentially exposed to SDP did not differ in their academic scores. The pattern of results suggests that genetic factors shared by parents and their offspring help explain why offspring exposed to SDP have lower levels of AA.

Maternal smoking during pregnancy (SDP) has been consistently linked with numerous birth complications and problems in the offspring, including unsuccessful pregnancy outcomes, lower intellectual abilities, lower academic achievement (AA), more inattention and hyperactivity problems, and increased antisocial behavior (reviews in Cnattingius, 2004; Huizink & Mulder, 2006; Wakschlag, Pickett, Cook, Benowitz, & Leventhal, 2002). Many researchers have noted that the current state of research suggests the associations are consistent with a causal connection as the findings are frequently consistent across studies, there is some specificity in the outcomes, the associations follow a dose–dependent relation, and basic research (e.g., animal studies) has identified plausible biological mechanisms through which SDP may interfere with development. The interest in exploring the effects of SDP stems from a growing appreciation of the role of prenatal risk factors for later development (e.g., Coe & Lubach, 2008) and the fact that SDP may be one of the most preventable risk factors for poor pregnancy outcomes (Cnattingius, 2004) and later antisocial behavior (Wakschlag et al., 2002).

It is important to note, however, that SDP is correlated with many risk factors that also predict poor outcomes in children. SDP is correlated with psychosocial risks, such as early age at childbearing; parental history of psychopathology and antisocial behavior, low socioeconomic status (SES), poor parenting, and ongoing exposure to parental substance use (reviews in Huizink & Mulder, 2006; Maughan, Taylor, Caspi, & Moffitt, 2004). These correlated risks could confound the association between SDP and offspring adjustment. A number of researchers have also argued that genetic confounds may explain the associations between SDP and offspring adjustment (D'Onofrio et al., 2003; D'Onofrio et al., 2008; Fergusson, 1999; Knopik et al., 2005; Silberg et al., 2003). Behavior genetic studies have documented that genetic factors influence maternal SDP (Agrawal et al., 2008; D'Onofrio et al., D'Onofrio et al., 2003). Genetic factors that influence SDP, such

---

Funding for the study was provided by grants from the Indiana University Faculty Research Support Program, NARSAD, and the Swedish Research Council. Preliminary analyses were presented at the Behavior Genetics Association Conference (June 2007, Amsterdam).

Correspondence concerning this article should be addressed to Brian D'Onofrio, Department of Psychological and Brain Sciences, Indiana University, 1101 E. 10th St., Bloomington, IN 47405. Electronic mail may be sent to bmdonofr@indiana.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8101-0006

as risk for substance abuse and dependence (Agrawal et al., 2008), could be passed from parents to their offspring and influence the behavior of their offspring. Thus, SDP may not be causally related to offspring behavior, but rather genetic factors shared by both generations could explain the association.

Approaches to Control for Selection Factors

Most studies of SDP have tried to control for selection factors that could lead to both SDP and offspring adjustment by statistically controlling for characteristics of mothers that covary with SDP (e.g., maternal education, SES, and history of delinquency). To date, reviews suggest that the associations between SDP and birth outcomes (Cnattingius, 2004), and antisocial behavior (Wakschlag et al., 2002) remain largely robust to the inclusion of statistical covariates. With respect to the association between SDP and low AA and intellectual abilities, the literature is less conclusive. Numerous studies have found an independent association between SDP and offspring AA when controlling for various familial characteristics (e.g., Butler & Goldstein, 1973; Cornelius, Ryan, Day, Godschmidt, & Willford, 2001; Fogelman & Manor, 1988; Rantakallio, 1983). In contrast, large national cohort (Fergusson & Lloyd, 1991) and nationally representative studies (Batty, Der, & Deary, 2006) suggest the association between SDP and academic problems is completely confounded by covariates, such as detailed measures of maternal intellectual abilities. Ultimately, the use of measured covariates is limited; finding an independent association while controlling for possible confounds can help support a possible causal inference, but it is impossible to know whether every salient confound was included and measured accurately in each study.

Because SDP co-occurs with numerous psychosocial and genetic risks for offspring the effects of SDP are difficult to isolate using standard family research designs. Quasi-experimental studies, approaches that disentangle co-occurring genetic and environmental risks, are therefore required to elucidate the processes through which SDP is associated with poor adjustment (Moffitt, 2005; Rutter, 2007). Because of the inherent limitations and strengths of each approach, converging evidence from multiple designs is needed to infer a causal association (Rutter, Pickles, Murray, & Eaves, 2001).

Researchers must use numerous approaches, including quasi-experimental designs, to study the putative effects of SDP. Recently, a number of genetically informed approaches have been used as quasi-experimental approaches. Two studies of SDP have utilized a Children of Twins design (D'Onofrio et al., 2003; Heath, Kendler, Eaves, & Markell, 1985; Silberg & Eaves, 2004), an approach that helps control for environmental factors that are not measured by comparing cousins differentially exposed to a putative risk factor. In this design environmental factors that influence all cousins (e.g., SES, religious affiliation) are controlled. The approach also accounts for genetic risk passed down from parents to their offspring because the offspring of an identical twin shares 25% of his or her genetic makeup with the offspring of his or her parent's cotwin (socially they are cousins, but genetically they are half siblings), whereas the offspring of a fraternal twin shares 12.5% of his or her genetic makeup with his or her cousin. D'Onofrio et al. (2003) found that the association between SDP and offspring birth weight (BW) was not due to environmental or genetic confounds. Rather, the association appeared to be mediated via environmental risks specifically associated with SDP, consistent with a causal inference. Knopik et al. (2005), however, found that the association between SDP and offspring attention deficit hyperactivity disorder (ADHD) was largely accounted for by genetic and environmental correlates of SDP, not the teratogenic effects of prenatal nicotine exposure.

A number of studies have recently utilized a sibling comparison approach, a design that compares siblings who are differentially exposed to SDP (the mother varied her smoking among her pregnancies). The comparison of siblings within families rules out all unmeasured environmental risks that affect siblings similarly (Rodgers, Cleveland, van den Oord, & Rowe, 2000). The comparison of siblings also controls for genetic factors transmitted from the mother that influence both SDP and AA in offspring (D'Onofrio et al., 2007; Harden et al., 2007; Lahey, D'Onofrio, & Waldman, 2009). This is because each child receives a random draw of their mother's genes through the process of meiosis, regardless of the mother's SDP during each pregnancy. It is also important to note that child effects, including the possibility of active or evocative gene–environment correlation, cannot confound the findings, as it is unlikely that traits in the offspring influence their mother's decision to smoke while pregnant (Rutter, 2007).

Recent use of the approach suggests that the direct effects of SDP on certain outcomes are minimal. D'Onofrio et al. (2008) found no association between SDP and offspring conduct and behavioral problems when comparing differentially exposed offspring. The association between SDP and attention/impulsivity problems was still statistically significant, but the magnitude of the association was greatly reduced when compared with the association among unrelated individuals. Of particular relevance to the current manuscript, Lambe, Hultman, Torrang, MacCabe, and Cnattingius (2006) found that among mothers who varied in their SDP across pregnancies, the risk for poor AA at age 15 was increased in all offspring regardless of whether they were specifically exposed to SDP. The study was a population-based study in Sweden created by merging national registries. The results suggested that the association between SDP and school performance at age 15 was due to unmeasured familial traits that differ between offspring exposed and unexposed to SDP.

The sibling comparison design is limited, however, in its ability to identify the source of the confounds when the risk factors are not associated with the outcome among siblings (D'Onofrio et al., 2008). Comparing siblings differentially exposed to SDP can disconfirm a causal hypothesis (e.g., Lambe et al., 2006), but the design cannot determine why there is a statistical relation between SDP and AA in the population if differentially exposed siblings do not also differ in AA. Researchers must, therefore, use additional, genetically informed approaches to explore whether factors that confound the association between SDP and offspring AA are genetic or environmental in origin. Numerous genetically informed approaches are available (Rutter et al., 2001).

One approach to disentangling the source of the confounds is to compare the magnitude of the association between SDP and offspring AA in groups of relatives who differ in their exposure to SDP and who differ in their genetic relatedness (D'Onofrio et al., 2008). When multiple types of relatives are included in a study (e.g., cousins, full siblings, and half siblings), researchers can examine whether the degree to which the groups of relatives share genetic risk moderates the association between SDP and AA (i.e., the degree to which relatives share genetic background influences how strongly SDP and AA are associated). If the association between SDP and AA is smaller when comparing groups of relatives who share more genetic background, then genetic factors would be implicated. Stated differently, if the magnitude of the association between SDP and AA is smaller when using comparison groups that better "control" for genetic risk (comparing groups of relatives who share more genetic background), the results would suggest that genetic factors help account for the statistical relation between SDP and AA that is found in the population.

When large family pedigrees are available, researchers can compare two types of cousins who are differentially exposed to SDP. Like children of fraternal twins, cousins whose mothers are full siblings share 12.5% of their genetic makeup (referred to as full cousins), whereas cousins of mothers who are half siblings provide a comparison group that shares only 6.25% of their genes (half cousins). Using full and half cousins, therefore, provides two groups of relatives who differ in the degree to which they share genetic risk; the comparison of full cousins would "control" for (or "account" for) more genetic risk than the comparison of half cousins.

Using different types of siblings can also help to differentiate the possible underlying mechanisms. The comparison of differentially exposed half siblings from the same mother (siblings who share 25% of their genetic makeup) only controls for genetic factors that are passed down from mothers. In contrast, the comparison of full sibling (offspring who share the same mothers and fathers) differentially exposed to SDP rules out genetic factors passed down from both mothers and fathers (Harden et al., 2007; Lahey et al., 2009). If genetic (or environmental) factors from fathers were important in explaining the association between SDP and AA, the association between SDP and AA would be smaller in full siblings than in half siblings (because the comparison of full siblings controls for more genetic risk). A difference in the associations between SDP and AA when comparing full siblings and half siblings would suggest that paternal effects due to assortative mating (the nonrandom mating of individuals based on similar backgrounds) are important in understanding the effects of SDP.

If the magnitude of the SDP and AA association decreases according to increasing genetic similarity in the groups of relatives (unrelated offspring > cousins sharing 6.25% of their genes > cousins sharing 12.5% of their genes > siblings who share 25% of their genes > siblings sharing 50% of their genes), genetic factors would be implicated. The findings would not prove that genetic factors account for the association because the various groups of relatives can also differ on environmental risk associated with SDP (see the Discussion). In contrast, the results would suggest the importance of environmental confounds (factors that influence both SDP and AA) if the association was of the same magnitude in all the relative groups. Environmental confounds would be implicated because the factors accounting for the association between SDP and AA would not depend on genetic risk associated with SDP. In sum, large family studies including relative groups who vary both in their exposure to SDP and in their genetic relatedness can help to differentiate whether genetic and/or environmental factors account for the statistical relation between SDP and AA.

The Current Study

The current study used a number of statistical and quasi-experimental approaches to explore the underlying causal mechanisms responsible for the association between maternal SDP and offspring AA. The current study expands on the study by Lambe et al. (2006), which compared siblings differentially exposed to SDP on school grades using offspring born in Sweden between 1983 and 1987. We use a larger sample to explore the association between SDP and grade average (using data on every individual born in Sweden between 1983 and 1991), include scores from a standardized assessment of mathematic proficiency given to all students in Grade 9 throughout the country, explore the possible confounding role of measured maternal and paternal characteristics, and utilize multiple quasi-experimental designs to help specify the causal mechanisms through which SDP is associated with lower AA.

There were four specific aims of the analyses: (a) identify the magnitude of the associations between SDP and two measures of AA, school grades and mathematics proficiency at age 15; (b) examine whether the relations were robust to the inclusion of statistical covariates for individual, maternal, and paternal characteristics; (c) explore the strength of the associations when comparing cousins and siblings differentially exposed to SDP; and (d) test whether genetic and/or environmental confounds help explain the statistical association between SDP and AA. The last aim was based on our ability to explore whether the strength of the statistical association between SDP and AA was dependent on the degree to which the exposed and unexposed individuals shared genetic risk. The analyses explicitly tested whether the association between SDP and AA was stronger when comparing cousins who share 6.25% of their genes (cousins of mothers who were half siblings) than when comparing cousins who share 12.5% of their genes (cousins of mothers who were full siblings). Likewise, the analyses explored whether SDP was more strongly associated with AA in half siblings (who share 25% of their genes) than in full siblings (who share 50%) who were differentially exposed to SDP.

Method

Sample

The current study was based on merging numerous Swedish population registries. Various governmental agencies and research institutes in Sweden maintain the registries. Each register includes a unique identifier for each individual in the country, which enables Statistics Sweden to link the various data sets. The registries and the agencies that maintain them are described in detail below (see Figure A1 in the Appendix for a diagram of the registries included).

The Medical Birth Register. The Swedish Medical Birth Register (MBR) is administered by the Swedish Centre for Epidemiology at the National Board of Health and Welfare. The MBR includes over 99% of all births in Sweden since 1973 (Swedish Centre for Epidemiology, 2001). Information is gathered throughout the pregnancy and delivery using standardized records. Since 1982, the register has also included information from the first visit to antenatal care. Over 95% of pregnant women in Sweden receive such care by the 15th gestational week (Lindmark & Cnattingius, 1991). The register includes detailed demographic and pregnancy information, including maternal age at birth, maternal living situation, maternal SDP, birth order of the child, infant physical attributes at birth, pregnancy complications, and delivery complications. No offspring who died at childbirth, had serious malformations at birth, and/or were missing the identification number of their mother were included. Because the current study explores prenatal risk factors, all multiple births (e.g., twins and triplets) were entirely excluded from the analyses. Children born in Sweden from 1983 to 1991 were included ($n = 918{,}028$).

The National School Register. The School Register, a collaboration between the Swedish School Authority and Statistics Sweden, includes educational achievement (scores in multiple subjects and summary grades) for all students at the end of Grade 9 (Swedish National Agency for Education, n.d.). The register began in 1983. The analyses in the current manuscript were conducted on information from students in all public and private schools throughout the country between 1998 and 2006. If students repeated the grade, the student's scores for his or her 1st year were used. The total School Register included grades for 972,979 individuals, including some individuals not included in the current analyses (e.g., those from multiple births and children who immigrated to Sweden).

Starting in 2003, all students in Grade 9 throughout Sweden completed a compulsory comprehensive test that provides standardized assessments of students' proficiency in Swedish, English, and mathematics. Thus, the standardized testing permits the comparison of students throughout Sweden on the same test. Scores were available for 456,007 students (from 2003 to 2006).

The National Crime Register. The National Crime Register is maintained by the Swedish National Council for Crime Prevention and includes all registered convictions in the country from 1973 to 2004. The register includes detailed information about the nature of every conviction, number of offenses, time of the criminal activity, and sentence type and length. Merging the data set in the study provided criminal background information from age 15 (age of criminal responsibility) for all of the mothers and fathers of the offspring born in Sweden from 1983 to 1991.

The Education Register. The Education Register is maintained by Statistics Sweden and includes information about the highest obtained level of education for each individual each year since 1990. We used the highest level of education in 2004 for the mothers and fathers of the target children in the study.

The 1990 Census. The Swedish Population and Housing Census were undertaken every 10 years until 1990. Its completion was enforced by law, and it included measures of individuals in each household, such as assessments of employment, occupation, and income. Occupational status information was obtained for individuals from the 1990 Census.

Multi-Generation Register. The Multi-Generation Register, maintained by Statistics Sweden, contains information on the identity of the parents of each child born in Sweden since 1932, or who immigrated and became Swedish citizens before age 18 (Statistics Sweden, 2003). The register is part of the Total Population Register, which collects information from the national registration records maintained by the National Tax Board. For each target individual, the register includes information on the identification number of biological parents and adoptive parents, if appropriate. The information allows researchers to link all children to their parents and identify all types of relationships for each index person (e.g., full siblings, half siblings, grandparents, cousins). Because some of our analyses comprise comparisons among cousins, offspring who had missing information about the identification of their maternal grandmother (e.g., their mother immigrated to Sweden from another country) were not included in the analyses.

Total sample and subsets included in analyses. The total sample used for analyses included 654,707 offspring with information on maternal SDP and at least one of the educational outcomes (either school grades from 1998 to 2006 or standardized testing from 2003 to 2006). The population included 469,182 individual mothers from 397,524 extended families (based on maternal lineage). See Figure A1 in the Appendix for more details.

The first subset of the data included unrelated offspring so that the between families effect of SDP on AA could be estimated. One grandchild from each extended family, indexed by the maternal grandmother, was initially selected. Offspring who shared any first-degree relative (e.g., their father) or second-degree relatives (e.g., their maternal grandfather, paternal grandmother, or grandfather) were excluded from the analyses. The highest genetic relatedness of any two offspring in the sample, therefore, would be minimal (second cousins, $r_g = .5^5 = .03125$). The unrelated subset included 301,823 offspring from the same number of mothers and maternal grandmothers.

The second subset of data was selected to compare cousins, especially cousins differentially exposed to mothers with a history of any SDP. Because the current study explores the putative effects of a maternal behavior, SDP, the identification of cousins was based on offspring of female sisters (i.e., each offspring in an extended family shared the same maternal grandmother). All offspring from extended families with two adult sisters were included in the analyses. Offspring of women with no adult sisters were excluded from the analyses because their offspring had no cousins who were related through maternal aunts. For extended families that contained more than two adult sisters, two adult women and their offspring were selected. If all the women in the extended family had similar histories of SDP (either never smoked during any pregnancy or smoked during every one) then the first two adult sisters and their families were included in the analyses. If there was any variability in SDP histories among the sisters (one woman had a history of ever SDP and one of her adult sisters did not) one adult sister with a history of SDP and one adult sister without and their offspring were included. The subset maximized, where possible, the number of cousins differentially exposed to mothers with histories of SDP. The subset, referred to as the cousin subset, included 181,219 offspring from 126,192 mothers and 63,096 maternal grandmothers.

The third subset of the data sought to compare siblings differentially exposed to SDP. The identification of siblings was based on sharing the same biological mother. All nuclear families that included only one offspring were excluded. All nuclear families with two children were included. If a woman had more than two children, two children were selected for inclusion in the subset. For mothers with no variation in their SDP across their pregnancies, her first two children were selected. For mothers who varied their SDP across their pregnancies, siblings were selected to include one offspring unexposed and one offspring exposed. The subset, referred to as the sibling subset, included 327,346 offspring from 163,673 mothers and 136,358 maternal grandmothers.

Variables

Offspring outcomes. School performance was based on a summary grade point score of 16 different subjects from the National School Register. Students were given separate grades (10 = pass, 15 = pass with distinction, and 20 = pass with honors) for each subject. The summary grade point score summed the subject scores for each student. As was found in previous studies (Lambe et al., 2006), the average grade point score in the population (M = 204.5, SD = 64.3, n = 761,970) was skewed toward higher values. Therefore, the variable was normalized using a Blom transformation, a procedure that has been shown to help when analyzing skewed data using genetically informed designs (van den Oord et al., 2000). The score was then converted to a z score for ease of interpretation.

The compulsory standardized testing of students in Grade 9 provided assessment of student proficiency in mathematics (n = 331,093). Students could receive four possible scores: goals not achieved (10.6%), pass with lowest pass grade (52.1%), pass with distinction or intermediate pass (26.4%), and pass with special distinction or highest pass grade (11.0%). The variable was transformed using a Blom transformation and standardized as a z score.

One birth characteristic, BW as measured in grams, was also included in the current study as a sensitivity analysis because previous findings strongly suggest SDP has a direct effect on offspring BW (e.g., Cnattingius, 2004). Offspring BW, excluding multiple births and children with significant physical malformations, was approximately normally distributed (M = 3,522.5 g, SD = 539.7, n = 759,627). Because BW was measured on an interpretable scale, no transformations were completed.

Risk factors. Maternal SDP was measured on a 3-point scale: nonsmokers, 1–9 cigarettes per day (moderate smokers), and 10+ cigarettes per day (heavy smokers). Given the complexity of the analyses, the analyses treated SDP as dichotomous (no smoking vs. smoking). As has been reported earlier (Cnattingius, 2004), the prevalence of SDP declined during the years of 1983–1992; women reported smoking during 31.3% of pregnancies in 1983 and 24.3% in 1992.

Risk factors specific to each pregnancy were also included. The MBR assessed maternal age at the time of birth, birth order, gestational age (in days), year of birth, and whether the mother was living with a partner at the time of birth, coded as not cohabiting versus cohabiting. Paternal age at the birth of the child was calculated based on his and the child's dates of birth.

Characteristics of the mothers and fathers were also included in the analyses. SES for both parents was measured by a frequently used, standard socioeconomic index variable, based on occupational group (Statistics Sweden, 1982). The groupings include (a) unskilled blue collar worker, (b) skilled blue collar worker, (c) low-level white collar worker, (d) intermediate-level white collar worker, (e) high-level white collar worker, (f) self-employed, (g) employed but uncategorized, and (h) not working/missing. The final category included individuals with no information about occupational status. It includes individuals too young to be assessed (below the age of 15) and individuals that were not working at the time of the census (because they were students or unemployed). Mothers and fathers' SES was based on the occupation of each individual separately. Previous research has shown that the groupings cannot be ordered (Lichtenstein, Pedersen, & McClearn, 1992). Therefore, dummy codes were created to compare each occupational group to the unskilled blue collar worker category.

Maternal and paternal highest level of education in 2004 was based on seven categories: (a) less than 9 years of primary and lower secondary education, (b) 9 years of primary and lower secondary education, (c) 1–2 years of upper secondary education, (d) 3 years of upper secondary education, (e) less than 3 years of postsecondary education, (f) more than 3 years of postsecondary education, and (g) postgraduate education. Because the assessment of educational level was not on an interval scale, dummy codes were used to compare higher levels of education to having completed less than 9 years of primary education. History of criminal behavior for mothers and fathers was based on the presence or absence of a conviction for any criminal offense in the National Crime Register through 2004.

Estimates of genetic relatedness. The biological mother of each offspring was determined by the MBR. The Multi-Generation Register provided information about the biological father of each child. From the information, two offspring from the same mother could be designated as full siblings (sharing the same biological father) or half siblings (having different biological fathers).

The determination of genetic relatedness of cousins was based on information from the Multi-Generation Register. The genetic relatedness of the cousins is based on the genetic correlations of adult sisters within an extended family (the mothers of the offspring in the current study). Genetic relatedness of the adult sisters in the study was determined contingent on their mothers, the maternal grandmothers of the offspring in the current study. Adult sisters who shared the same fathers (the maternal grandfather of the offspring in the study) were determined to be full siblings (sharing 50% of their genetic makeup), whereas adult sisters with different fathers were designated as half siblings (sharing 25% of their genetic makeup). Thus, offspring of full siblings, referred to as full cousins, shared 12.5% of their genes with their cousins, whereas offspring of half siblings, referred to as half cousins, shared 6.25% of their genes with their cousins. The process of determining the relationship of adult sisters contingent on the maternal grandmother does not include aunts who are half siblings because they had the same father but were born to different mothers. We restricted the sample to only include aunts who shared the same mother because the analyses explored the risk of a maternal characteristic (SDP) and we wanted to control for environmental factors that may be shared by women growing up in the same household; most children historically have lived with their mothers after separations or divorces.

Analyses

Association between SDP and risk factors. Linear and logistic regression models estimated the association between SDP and the risk factors at the individual, maternal, and paternal levels. Clustering for extended family members was taken into account using a sandwich estimator and full information maximum likelihood (review in Schafer & Graham, 2002) was utilized to account for missing data using the software program Mplus (Muthén & Muthén, 1998–2007).

Comparison of offspring exposed and unexposed to SDP contingent on biological relatedness. To describe the association between SDP and the AA variables and the underlying processes, the mean AA scores are presented for offspring differentially exposed using the three subsets of the data: unrelated offspring, cousins, and siblings. The difference between the AA means suggest the magnitude of the association between SDP and each academic measure. Formal significance testing of the differences was completed using regression analyses (see the next section). The key to understanding the underlying causal mechanisms associated with SDP stems from the difference in exposure, not the overall mean levels, which are heavily influenced by family type. The means for AA among offspring differentially exposed (e.g., cousins and siblings who varied in their exposure) are tabled.

The first comparison involved unrelated offspring who were exposed and unexposed to SDP. The difference between those exposed and unexposed represents the association between SDP and each AA variable in standard deviation units (the AA variables were converted to z scores). The comparison of cousins differentially exposed to mothers with a history of SDP, regardless of genetic relatedness, was then completed. The comparison controlled for factors influencing all offspring within an extended family. Next, the cousin comparison was estimated separately for full cousins (those sharing 12.5% of their genes) and half cousins (sharing 6.25% of their genes) to examine whether the genetic relatedness of the cousins moderated the association between SDP and AA.

Mean comparisons were then estimated for differentially exposed siblings, regardless of genetic relatedness. Because the identification of siblings was based on sharing the same mother in the current sample, the comparison controls for genetic factors passed from the mother to her offspring, as well as environmental factors shared by siblings. Next, the sibling comparison was calculated separately for full siblings (those who share 50% of their genes) and half siblings (those only sharing 25% of their genetic makeup) to examine the possible influence of paternal factors.

Regression analyses of the relations between SDP and offspring AA. A series of regression analyses were performed to formally test the associations between SDP and AA across the groups of relatives and include the statistical covariates. Five regression models were fit for school grades and math proficiency. Descriptions of the models, including the main parameters of interest, an explanation of the models, and the samples utilized, are found in Table 1. All models accounted for the clustering of the data and used full information maximum likelihood to handle missing values using Mplus (Muthén & Muthén, 1998–2007). Model 1, an epidemiological model, examined the magnitude of the association between SDP and school grades in the entire sample, controlling for offspring sex. Model 2, the standard approach for controlling for selection factors, estimated the association of SDP and grades in the full sample while statistically controlling for the measured covariates.

Model 3 examined the association when comparing cousins and siblings differentially exposed

Table 1
Description of Analytical Models

Description	Parameters of interest	Explanation	Sample used and controls
1. Epidemiological model	SDP parameter	Estimates the difference in AA between unrelated offspring exposed and unexposed to SDP	Includes all offspring in entire data set
2. Association controlling for statistical covariates	Compare magnitude of SDP parameter to Model 1	Estimates the independent association between SDP and AA while statistically controlling for covariates	Includes all offspring in the entire data set
3. Cousin and sibling comparison	SDP: cousin comparison SDP: sibling comparison	Estimates whether offspring whose mother had more SDP across her pregnancies have lower AA than their cousins whose mothers had less SDP on average Estimates whether siblings exposed to SDP had lower AA than their siblings who were not exposed	Includes all offspring in entire data set while controlling offspring birth order
4. Comparison of full and half cousins	SDP: half cousins SDP: half cousin–full cousin	Estimates the association between SDP and AA when comparing half cousins (share 6.25% of genetic makeup) who were differentially exposed Estimates the difference in the cousin SDP–AA association between half and full cousins, which was based on the interaction between the cousin parameter and cousin genetic relatedness, coded 0 = half cousin and 1 = full cousin. To calculate the magnitude of the SDP–AA association in full cousins requires adding the two cousin parameters in the model	Includes all offspring from two mothers per extended family while controlling for the main effects of cousin relatedness and offspring birth order
5. Comparison of full and half siblings	SDP: half siblings SDP: half sibling–full sibling	Estimates the association between SDP and AA when comparing half siblings (share 25% of genetic makeup) who were differentially exposed Estimates the difference in the cousin SDP–AA association between half and full siblings, which was based on the interaction between the sibling parameter and sibling genetic relatedness, coded 0 = half sibling and 1 = full sibling. To calculate the magnitude of the SDP–AA association in full siblings requires adding the two sibling parameters in the model	Includes two offspring from all mothers with two or more offspring while controlling for main effects of sibling relatedness and offspring birth order

Note. All models accounted for the clustering in the data using a sandwich estimator in Mplus. Each model also included the main effect of offspring sex. SDP = smoking during pregnancy; AA = academic achievement.

in the full sample while controlling for birth order effects. The models used a series of contrast codes that have been shown to accurately estimate within and between family effects at multiple levels (Neuhaus & McCulloch, 2006). The statistical approach results in parameters equivalent to that of a fixed effects model (Greene, 2003). For further details concerning the analytical approach, including explanations of the coding, see D'Onofrio et al. (2008) and Harden et al. (2007). Sibling comparisons were based on contrast codes created by subtracting the mother's average SDP across all of her pregnancies from the child's specific exposure. For example, if a mother had two children and only smoked during one pregnancy, the exposed sibling would receive a score equal to .50 and the unexposed sibling would receive a contrast code of −.50. The contrast codes, therefore, compare exposure to SDP relative to their sibling's exposure. A similar approach was used to compare cousins. Contrast codes were calculated by subtracting the average level of SDP across all women in an extended family (the adult sisters) from each mother's average SDP across all of her pregnancies. The contrast codes compared siblings from a nuclear family who were more exposed to SDP on average to their cousins who were less exposed to SDP on average. The mean SDP across all cousins was also included in the model because the parameter associated with the variable provides a comparison of unrelated individuals (i.e., the parameter compares cousins who are all exposed to more SDP on average to unrelated sets of cousins who were exposed to less SDP on average). Model 3, therefore, provides estimates of the association between SDP and AA when comparing unrelated individuals, cousins, and siblings.

The next two models were conducted to further explore the cousin comparisons (Model 4) and the sibling comparisons (Model 5). Model 4 specifically examined whether the magnitude of the cousin comparison depended on whether the cousins were full or half cousins. The model was fit to a subset of the data that only included two adult sisters and their offspring from each extended family so that only one cousin comparison could be made per extended family (a conservative approach to handling extended families with more than two adult sisters). The model utilized the contrast coding scheme for cousin comparisons explained above (recalculated for the subset) and the interaction of the contrast coding with a variable that identified whether the cousins were half (coded = 0) or full cousins (coded = 1). The interaction term, therefore, estimated the difference in the magnitudes of the associations between SDP and school grades in the two types of cousin pairs. The parameter associated with the interaction specifically estimated whether cousin type (full vs. half) moderates the association between SDP and AA, providing a statistical test of whether the SDP–AA association was stronger when comparing half cousins than when comparing full cousins (e.g., D'Onofrio et al., 2005). The model also included the main effect of cousin relatedness.

Model 5 explored whether the strength of the association between SDP and school grades varied among full and half siblings. The model was fit to the subset of data that only included two offspring per nuclear family, which permitted one sibling comparison per mother. The sibling comparison contrast code approach described earlier was used (recalculated on the subset). Plus, the interaction of the contrast code with a variable noting the genetic relatedness of the siblings (0 = half siblings, 1 = full siblings) was included in the model (as well as the main effect of sibling type). The parameter associated with the interaction term assessed whether the SDP–AA association differed between full and half siblings.

Comparison of fathers of half siblings differentially exposed to SDP. Studying nuclear families that include half siblings (the siblings had different fathers) who were differentially exposed to SDP provides a unique opportunity to examine risk factors associated with SDP. By comparing the characteristics of the two fathers and the pregnancies that are associated with the mothers SDP allows us to examine whether any pregnancy-specific or paternal risk factors were specifically associated with the pregnancy during which the mother smoked. Therefore, regression models predicting prenatal smoking exposure were fit to a sample of offspring (n = 2,773) from families that included siblings with different biological fathers and differential exposure to SDP.

Sensitivity analyses. A series of analyses examined the robustness of the findings presented in the manuscript and are briefly summarized in the text (full results available upon request).

Results

Association Between SDP and Risk Factors

Table 2 presents the relation between maternal SDP and each risk factor. The associations indicate

Table 2

Association Between Maternal Smoking During Pregnancy and Child-Specific, Maternal, Paternal, and Familial Risks

Risk factor	*b*/OR	Confidence interval	*n*
Child-specific risks[a]			
Maternal age at childbirth (years)	*b* = −1.17***	−1.20 to −1.14	654,707
Paternal age at childbirth (years)	*b* = −.95***	−0.99 to −0.92	654,707
Birth order	*b* = .014***	0.01 to 0.02	654,707
Gestational age (days)	*b* = −1.15***	−1.22 to −1.08	654,707
Year of birth (years)	*b* = −.25***	−0.03 to −0.24	654,707
Not cohabiting	OR = 3.58***	3.46 to 3.71	654,707
Maternal risks[b]			
Occupation			
Unskilled blue collar	Ref	Ref	181,772
Skilled blue collar	1.00	0.98 to 1.01	76,421
Low-level white collar	OR = 0.89***	0.88 to 0.91	94,642
Intermediate-level white collar	OR = 0.39***	0.38 to 0.40	114,688
High-level white collar	OR = 0.36***	0.34 to 0.37	40,467
Self-employed	OR = 0.78***	0.75 to 0.81	16,563
Employed, uncategorized	OR = 1.46***	1.44 to 1.48	43,791
Not working/missing	OR = 1.39***	1.36 to 1.41	85,262
Educational level			
Less than 9 years of education	Ref	Ref	7,751
9 years of education	OR = 2.69***	2.64 to 2.75	68,264
1–2 years upper secondary education	OR = 1.67***	1.65 to 1.69	250,267
3 years upper secondary education	OR = 0.88***	0.86 to 0.89	89,323
Less than 3 years postsecondary education	OR = 0.48***	0.47 to 0.49	112,728
3+ years postsecondary education	OR = 0.37***	0.36 to 0.38	115,699
Postgraduate education	OR = 0.21***	0.18 to 0.24	3,091
History of criminal conviction	OR = 2.58***	2.54 to 2.63	469,182
Paternal risks[b]			
Occupation			
Unskilled blue collar	Ref	Ref	135,529
Skilled blue collar	OR = 1.34***	1.32 to 1.36	147,089
Low-level white collar	OR = 0.81***	0.79 to 0.83	55,462
Intermediate-level white collar	OR = 0.56***	0.55 to 0.58	103,552
High-level white collar	OR = 0.38***	0.37 to 0.39	87,224
Self-employed	OR = 0.91***	0.89 to 0.93	46,690
Employed, uncategorized	OR = 1.80***	1.77 to 1.83	30,128
Not working/missing	OR = 1.82***	1.78 to 1.86	42,507
Educational level			
Less than 9 years of education	Ref	Ref	28,812
9 years of education	OR = 1.77***	1.74 to 1.80	101,413
1–2 years upper secondary education	OR = 1.52***	1.50 to 1.54	240,474
3 years upper secondary education	OR = 0.81***	0.80 to 0.83	77,191
Less than 3 years postsecondary education	OR = 0.55***	0.54 to 0.56	83,455
3+ years postsecondary education	OR = 0.36***	0.35 to 0.36	89,417
Postgraduate education	OR = 0.24***	0.22 to 0.26	8,946
History of criminal conviction	OR = 2.22***	2.20 to 2.25	469,182

Note. All estimates accounted for clustering using a sandwich estimator and were based on full maximum likelihood approaches to account for missing data. OR = odds ratio.

[a] Estimates for the child-specific risks are based on maternal smoking during each specific pregnancy for all births.

[b] Estimates are unique for each individual mother and are based on maternal history of ever smoking during any pregnancy. The associations with paternal characteristics are based on the father of the first child per mother.

***$p < .001$.

that maternal SDP is associated with the covariates at all levels. When exploring each pregnancy, SDP was associated with earlier age at childbearing for parents, shorter gestational age, and a higher likelihood that the mother was not living with a partner at the time of birth.

The association between SDP and maternal, paternal, and familial risks were based on a woman's history of ever smoking across her pregnancies. SDP was associated with maternal occupational status; compared to women in unskilled blue collar jobs, SDP was associated with lower likelihood of working in white collar jobs. SDP was also associated with greater likelihood of being in jobs that were difficult to categorize or being unemployed. SDP was also associated with maternal highest level of educational achievement; women with a history of SDP were less likely to receive secondary and higher levels of education. Being convicted of a criminal offense was also associated with history of SDP. The same general pattern held for paternal risks. SDP was associated with occupational status, highest educational level achieved, and history of criminal offenses in fathers.

Comparison of Offspring Exposed and Unexposed to SDP Contingent on Biological Relatedness

Table 3 presents the mean comparisons of offspring differentially exposed, based on family type. The mean AA scores for offspring from extended and nuclear families with no variation in SDP are

Table 3

Mean Offspring Outcomes Based on Exposure to Maternal Smoking During Pregnancy and Genetic Relatedness to Comparison Group

	Grades		Math scores	
SDP	*M*	*n*	*M*	*n*
Unrelated offspring[a]				
Unexposed	0.19	210,858	0.14	97,286
Exposed	−0.30	79,954	−0.24	31,417
Difference (*SD*)	0.49		0.38	
Cousin comparison (all cousins)[b]				
No history SDP	−0.07	30,656	−0.04	13,618
History of SDP	−0.32	31,273	−0.24	12,440
Difference (*SD*)	0.25		0.20	
Cousin comparison (aunts were full siblings, $r_g = .125$)[b]				
No history SDP	−0.04	26,524	−0.02	11,839
History of SDP	−0.28	26,794	−0.22	10,667
Difference (*SD*)	0.24		0.20	
Cousin comparison (aunts were half siblings, $r_g = .0625$)[b]				
No history SDP	−0.23	3,313	−0.14	1,428
History of SDP	−0.55	3,351	−0.41	1,322
Difference (*SD*)	0.32		0.27	
Sibling comparison (all siblings)[c]				
Unexposed	−0.19	16,467	−0.16	8,117
Exposed	−0.16	16,465	−0.18	5,423
Difference (*SD*)	0.03		0.02	
Sibling comparison (full siblings, $r_g = .50$)[c]				
Unexposed	−0.16	14,675	−0.15	7,356
Exposed	−0.13	14,579	−0.14	4,801
Difference (*SD*)	0.03		0.02	
Sibling comparison (half siblings, $r_g = .25$)[c]				
Unexposed	−0.54	1,377	−0.34	583
Exposed	−0.63	1,396	−0.51	525
Difference (*SD*)	0.09		0.17	

Note. SDP = smoking during pregnancy.

[a] Based on one grandchild per maternal grandmother. No offspring shared any other first or second degree relatives in the subset.

[b] Based on cousins from two aunts from each extended family.

[c] Based on two siblings for each mother. Significance testing of the differences were conducted using regression analyses controlling for clustering in the data and birth order (see Tables 4 and 5).

presented in Table A1 (in the Appendix). The comparison of unrelated offspring suggests a 0.49 *SD* difference in grades associated with exposure to SDP. The comparison of all cousins differentially exposed to mothers with a history of SDP, regardless of genetic relatedness, suggests the association is attenuated (compared to the study of unrelated individuals), as the difference between the exposed and unexposed was approximately 0.25 *SD*. The comparison of full cousins was comparable (0.24 *SD* difference), but the comparison of half cousins yielded a larger difference (0.32 *SD*), suggesting that the degree of genetic relatedness moderates the association.

The comparison of differentially exposed siblings suggests SDP is not associated with lower grades. When all siblings, regardless of genetic relatedness, are compared there was no difference (exposure is associated with a 0.03 *SD* increase in grades). The comparison of full siblings revealed a similar difference (0.03 *SD* increase). In contrast, the comparison of half siblings (−0.09 *SD*) suggests that SDP is associated with offspring grades when comparing siblings with different fathers.

The same pattern of results was true when comparing the math proficiency scores. The comparison of unrelated individuals revealed a large difference (0.38 *SD*). The comparison of cousins revealed an attenuated difference that was partially dependent on genetic relatedness (a 0.20 difference in full cousins but a 0.27 difference in half cousins). There was no difference in math proficiency when all siblings were compared (0.02 difference). Similarly, no difference emerged when full siblings were compared (0.02 increase), but the exposure to SDP was associated with poorer math proficiency (0.17 *SD* difference) in half siblings.

The mean comparisons suggested two main findings. First, for both school grades and math proficiency, the comparison of full siblings differentially exposed to SDP, a comparison that controls for genetic and environmental confounds shared by siblings, revealed no differences. This finding strongly suggests that SDP does not cause lower AA; rather, differences between families associated with SDP appear to be responsible. Second, the association between SDP and AA appears to covary with degree of genetic relatedness. Table 3 illustrates how the difference between the exposed and unexposed offspring depended on genetic relatedness. For both measures of AA, the magnitude of the association between SDP and AA declined as genetic relatedness increased, suggesting that genetic factors help to explain the statistical relation between SDP and AA.

Regression Analyses of the Relations Between SDP and Offspring AA

School grades. The results of the regression models for school grades are presented in Table 4. Model 1, an epidemiological model, indicated SDP is associated with a 0.50 *SD* decrease in school grades (given the large sample size, significance values will only be noted when they are not $p < .05$). Model 2, which statistically controlled for the covariates, indicated that SDP was associated with a 0.21 *SD* decrease, independent of the effects of the other covariates in the model (Table A2 in the Appendix includes the parameter estimates for all of the covariates). The results from Model 3, a model that compared all cousins and siblings differentially exposed, show that SDP is associated with lower school grades when comparing unrelated offspring ($b = -.58$). The association is attenuated when comparing cousins ($b = -.26$) but is nonexistent when comparing siblings ($b = .00$, $p = .82$).

The results of the Model 4, a model that assessed whether cousin type moderated the association between SDP and AA, suggest that SDP exposure is more strongly associated with school grades when comparing half cousins ($b = -.31$) than when comparing full cousins ($b = -.23 = -.31 + .08$). It is important to note that the difference ($b = .08$) was statistically significant, indicating that cousin type moderates the association between SDP and school grades.

Model 5 examined whether the association between SDP and AA was stronger in half siblings than in full siblings. The model indicated that the half siblings who were differentially exposed to SDP differed in their school grades ($b = -.08$), whereas full siblings differentially exposed did not differ ($b = .00 = -.08 + .08$). The difference was statistically significant.

Figure 1, Panel A presents the estimates of the association between SDP and school grades separately for the different comparison groups (unrelated offspring, half cousins, full cousins, half siblings, and full siblings). The figure illustrates how the association between SDP and school grades is dependent on the genetic relatedness of the relatives used in the comparisons.

Mathematics proficiency. Model 1 indicated that offspring exposed to SDP performed less well on a standardized mathematics test ($b = -.39$ *SD*). The

Table 4

Regression Analyses of the Association Between Maternal Smoking During Pregnancy and Offspring Grade Average

	Model 1		Model 2^a		Model 3		Model 4		Model 5	
Risk factor	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*
Smoking during pregnancy										
Unrelated comparison	−.50	0.00	−.21	0.00	−.58	0.00	−.67	0.01	−.56	0.00
Cousin comparison										
All cousins					−.26	0.01				
Half cousins							−.31	0.02		
Full cousin–half cousin							.08	0.02		
Sibling comparison										
All siblings					.00*	0.01				
Half siblings									−.08	0.03
Full sibling–half sibling									.08	0.03
Offspring sex	.39	0.00	.39	0.00	.39	0.00	.39	0.00	.39	0.00
Birth order			−.14	0.00	−.12	0.00	−.10	0.00	−.10	0.00
Aunts are full siblings							.24	0.01		
Siblings are full siblings									.45	0.01
Number of offspring	654,707		654,707		654,707		181,099		327,161	

Note. See Table 1 for a description of all of the models.
^{a}Parameter estimates for all of the covariates are presented in the Appendix.
$^*p > .05$.

results are presented in Table 5. The association ($b = -.15$) was attenuated when statistically controlling for measured risk factors in Model 2 (see Table A3 in the Appendix). Model 3 indicated that the association between SDP and math scores was smaller when comparing cousins ($b = -.20$) than unrelated individuals, and there was no association when siblings were compared ($b = .00$, $p = .94$). Model 4, which explored whether the magnitude of the cousin comparison was dependent on the type of cousins suggested that the association between SDP and mathematics proficiency was stronger when comparing half cousins ($b = -.25$) than full cousins ($b = -.19 = -.25 + .06$), although the difference between the two cousin types could not be estimated precisely ($p = .14$). Model 5, however, indicated that the association between SDP and math scores is stronger (and statistically different) in half siblings ($b = -.15$) than full siblings ($b = .01 = -.15 + .16$). Figure 1, Panel B shows how the association between SDP and math scores varies across the types of relatives.

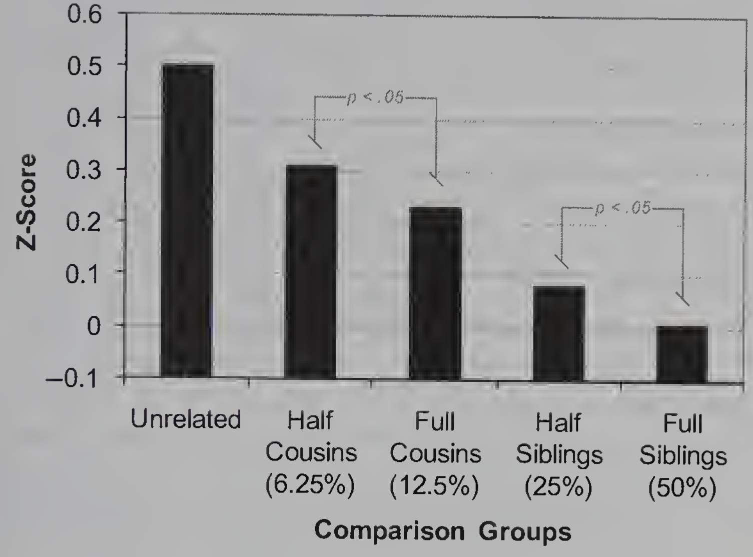
Comparison of Fathers of Half Siblings Differentially Exposed to SDP

The model fitting described earlier suggest that SDP is associated with slightly poorer grades and math scores when comparing siblings who have different fathers. As half and full siblings are equally genetically related to their mothers, the differences in association between SDP and AA for half and full siblings may be due to a genetic correlation between fathers and mothers. A series of regression analyses were conducted to examine which, if any, child-specific and paternal risks factors were associated with the pregnancy during which the mother smoked when the siblings were (a) differentially exposed to SDP and (b) had different fathers ($n = 2{,}773$ births). Birth order, maternal and parental age at childbearing, paternal education, and paternal occupational status were not associated with the pregnancy during which the mother smoked. However, the father of the child who was exposed to SDP was more likely to have a history of criminality (odds ratio [OR] = 1.40, confidence interval [CI] = 1.19–1.62, $p < .05$) and the mother was more likely not to be living with any partner at the time of the birth (OR = 2.05, CI = 1.60–2.63, $p < .05$).

Sensitivity Analyses

A number of additional analyses were performed to test the validity of the findings. Analyses of BW were completed to replicate previous research findings. When comparing unrelated individuals, SDP was associated with a 205-g decrease in BW. The comparison of siblings differentially exposed indicated SDP was associated with a 101-g

A. Grade 9 School Average



B. Grade 9 Mathematics Proficiency

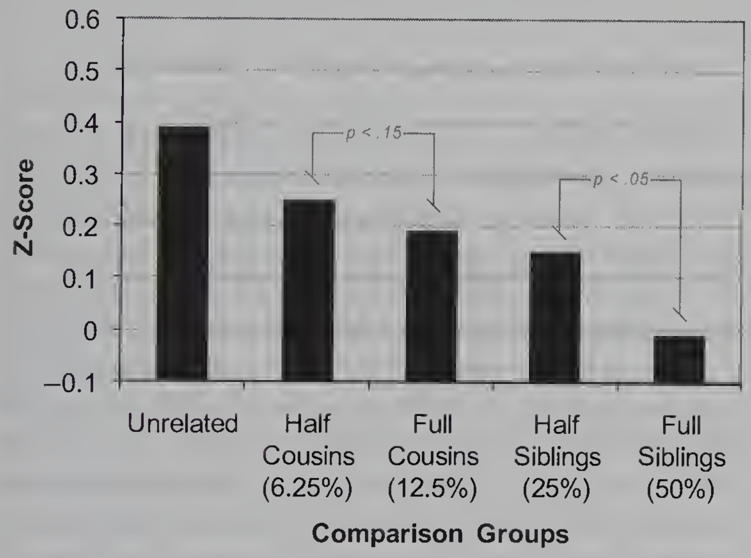


Figure 1. Associations between smoking during pregnancy and offspring academic achievement based on degree of genetic relatedness.
Note. Statistical tests of the differences among the cousin and sibling groups are based on Models 4 and 5, respectively.

decrease. The difference in magnitude of the effect when comparing half and full siblings (20 g) was not statistically significant (p = .32). As no moderation by genetic relatedness was found, the results strongly suggest that SDP has a specific effect on offspring BW.

Results of regression models predicting dichotomous measure AA, a measure of school grades (below passing) using the same cutoff as previous studies (Lambe et al., 2006) and predicting failing mathematics (goals not achieved), were comparable to those reported in the current manuscript (data not shown). Results of analyses using dummy codes to compare moderate smoking and heavy smoking to nonsmokers revealed comparable findings to the results of considering SDP as present versus absent.

Discussion

The current study combined two general quasi-experimental approaches to study the mechanisms through which SDP influences offspring AA. The study relied on the historical decline in SDP, due in part to public health campaigns to deter pregnant women and women of childbearing ages from smoking (Cnattingius, 2004), to explore whether variations in SDP by the same mother were associated with differences in her offspring's AA. The study also utilized a large family study, including a mixture of sibling and cousin types, to compare relatives who differ in their exposure to SDP and who vary in the degree to which they share genetic risk associated with SDP. By combining these two approaches, the current study sought to pull apart the co-occurring genetic and environmental risks associated with SDP.

The results provide support for two main conclusions. First and foremost, the results strongly suggest that SDP does not cause offspring to have lower AA. We compared full siblings who were differentially exposed to SDP, an approach that accounts for the genetic factors and environmental factors that siblings share (Lahey et al., 2009; Rutter, 2007). If SDP causes offspring to have lower AA, perhaps through the neurotoxic effects of exposure to prenatal nicotine, then a child exposed to SDP would be expected to have poorer AA than his or her sibling who was not exposed. For both school grades and mathematics proficiency, however, full siblings differentially exposed to SDP did not differ in their levels of AA. The results strongly suggest that familial risks correlated with SDP, and not the specific effects of SDP, are responsible for the lower AA found in offspring whose mothers smoked during pregnancy. The findings are consistent with previous research on the role of SDP and AA in Sweden (Lambe et al., 2006), as well as a sibling comparison study of conduct problems in the United States (D'Onofrio et al., 2008) and a Children of Twins study of ADHD in Australia (Knopik et al., 2005). The results are also consistent with traditional family studies that used extensive covariates (Batty et al., 2006; Fergusson & Lloyd, 1991). These designs, however, do not explain why SDP is associated with lower levels of AA in the population.

Table 5
Regression Analyses of the Association Between Maternal Smoking During Pregnancy and Offspring Grade 9 Mathematics Proficiency

	Model 1		Model 2^{a}		Model 3		Model 4		Model 5	
Risk factor	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*
Smoking during pregnancy										
Unrelated comparison	–.39	0.01	–.15	0.01	–.45	0.00	–.52	0.00	–.44	0.01
Cousin comparison										
All cousins					–.20	0.01				
Half cousins							–.25	0.04		
Full cousin–half cousin							.06*	0.03		
Sibling comparison										
All siblings					.00*	0.02				
Half siblings									–.15	0.06
Full sibling–half sibling									.16	0.06
Offspring sex	.02	0.00	.03	0.00	.03	0.00	.03	0.00	.03	0.00
Birth order			–.11	0.00	–.08	0.00	–.08	0.00	–.10	0.00
Aunts are full siblings							.18	0.01		
Siblings are full siblings									.30	0.01
Number of offspring	283,938		283,938		283,938		77,622		142,559	

Note. See Table 1 for a description of all of the models.
^{a}Parameter estimates for all of the covariates are presented in the Appendix.
$^{*}p < .05$.

The second conclusion involves the underlying processes by which SDP is associated with AA. The results suggest that genetic factors passed from parents to their offspring (passive gene–environment correlation) account for at least part of the statistical association. Genetic factors are implicated because the degree to which relatives were genetically correlated moderated the association between SDP and AA when comparing types of cousins and types of siblings. Models 4 and 5, in fact, provided statistical tests of these differences. When relatives were more genetically similar (full cousins vs. half cousins and full siblings vs. half siblings), the association between SDP and AA was weaker. Stated differently, when genetic factors were more controlled (by comparing relatives who shared more genetic factors), the association between SDP and AA went down. Figure 1 graphically illustrates this finding. It is important to note that the findings were consistent across levels of analysis (i.e., the comparisons of half vs. full cousins were similar to those comparing full and half siblings) and measures of AA, which provides converging evidence for the confounding role of genetic factors.

We stress that the results do not prove that the underlying causal mechanisms are genetic—the findings only suggest that genetic factors confound the association (i.e., the results are consistent with the role of genetic confounds). Half cousins and full cousins differ in their genetic relatedness, but they also may differ on environmental risks. Full cousins may see each other (and their aunts) more often than half cousins, which could make full cousins more similar if the level of contact influenced AA—the equal environments assumption one generation removed (D'Onofrio et al., 2003). Offspring of half siblings may also be exposed to more environmental risk than offspring of full siblings, which could influence the variability in AA. Similarly, there may be differences in the variability of AA between half and full siblings. These limitations, which apply to all step-family behavioral genetic designs (e.g., Reiss, Neiderhiser, Hetherington, & Plomin, 2000) need to be tested in future studies. Using the Children of Twins design would help test these assumptions. Studies of adult twins frequently include measures of contact between the twins (and their families), which could help account for differences in levels of family contact. Differences in genetic risk among cousins also would not be based on comparing offspring from intact families to those from divorced or separated parents; rather, differences in genetic relatedness among cousins in the Children of Twins design are due to the zygosity type of the adult twins.

The comparison of fathers of half siblings discordant for exposure to prenatal smoking indicated that the father of the child exposed to SDP more often had a history of criminal convictions. Exposure to prenatal smoking, thus, covaries with paternal risk of criminality, even within families. For the paternal phenotype to influence a genetic association between mother and offspring, the paternal phenotype has to be genetically correlated with the maternal phenotype. It is well known that there is substantial assortative mating for antisocial behavior (e.g., Krueger, Moffitt, Caspi, Bleske, & Silva, 1998), and in the current sample women with a history of a criminal conviction were much more likely to have children with men with convictions (OR = 3.38, CI = 3.32–3.46). Thus, the pattern of associations between SDP and AA for half and full siblings is congruent with genetic confounding, even though environmental explanations cannot be excluded. Additional research is required to further delineate the underlying causal mechanisms.

The fact that SDP is associated with lower AA in half siblings provides compelling evidence that researchers must explore the role of fathers when studying the putative effects of SDP. Although most research on SDP has not included measured characteristics of the fathers (review in Maughan et al., 2004), the current results indicate that characteristics of fathers, whether due to genetic risk passed down to the children or psychosocial risk factors, help explain some of the association between SDP and offspring AA.

In addition to the ability to use quasi-experimental approaches to study the risks associated with SDP, the current study also benefits from a number of key strengths. The study was conducted on a large, national sample of offspring. The analyses also included two measures of AA, school grades and mathematics proficiency. The inclusion of the mathematics scores greatly helps the interpretation of the data because the assessment was standardized across the entire study. School grades could vary substantially by schools or communities. The fact that both measures of AA provide similar results provides converging evidence for the conclusions drawn. The large sample size also permits the comparison of unique types of relatives (e.g., half siblings differentially exposed to SDP) that would be hard to find using traditional samples. SDP was also assessed during pregnancy rather than relying on maternal retrospective reports, a key limitation of some of the previous sibling comparison tests of SDP (Rutter, 2007). The analyses were also able to include measured covariates of both mothers and fathers. The exclusion of paternal information is a key limitation of previous research on SDP (Maughan et al., 2004).

There are also a number of limitations of the current study. The measure of SDP was based on maternal report, usually during the beginning of her pregnancy. Exposure to smoking later in life was also not available. Therefore, the study cannot explore whether the timing of the exposure was critical for later AA. The study could also not explore the importance of paternal SDP. The analyses also did not study whether individual or familial factors moderate the association between SDP and offspring AA. Research has suggested that factors, such as offspring sex and birth complications (Brennan, Grekin, & Mednick, 1999), interact with SDP to substantially increase risk for poor offspring adjustment. The comparison of siblings differentially exposed relies on assumptions that mothers who vary their smoking across pregnancies are comparable to those who smoked during every pregnancy (D'Onofrio et al., 2008). Furthermore, the analyses did not estimate the degree to which environmental and genetic confounds explain the association between SDP and AA. The results imply that genetic factors (and environmental confounds) are important. We are currently working on the analytical models required to specifically quantify the role of genetic and environmental factors when studying exposure to SDP with the various types of relative pairs available. More research, therefore, is needed to understand the causal process responsible for the association between SDP and poor offspring AA.

Overall, the study illustrates the importance of quasi-experimental approaches to studying putative environmental risk factors (Moffitt, 2005; Rutter, 2007). Similar to the two previous sibling comparison studies of SDP (D'Onofrio et al., 2008; Lambe et al., 2006), we would have drawn the wrong conclusion about the role of SDP if we only relied on the statistical covariates to account for confounds. Even though the regression analyses on the epidemiological sample controlled for measures of maternal and paternal traits, we would have wrongly concluded that SDP had an independent association with AA after controlling for maternal and paternal characteristics, consistent with a causal inference.

Finally, it is important to emphasize that the current results do not suggest that SDP has no impact on offspring adjustment. The results are limited to measures of AA at age 15 only. The findings for BW suggest, as have previous genetically informed (D'Onofrio et al., 2003) and other quasi-experimental studies (Cnattingius, 2004), that SDP has a specific effect on offspring BW that is environmentally mediated. SDP appears to impact particular outcomes more strongly than others, particularly poor pregnancy outcomes (Cnattingius, 2004). Reducing SDP, therefore, remains a major public health issue.

References

Agrawal, A., Knopik, V. S., Pergadia, M. L., Waldron, M., Bucholz, K. K., Martin, N. G., et al. (2008). Correlates of cigarette smoking during pregnancy and its genetic and environmental overlap with nicotine dependence. *Nicotine and Tobacco Research*, *10*, 567–578.

Batty, G. D., Der, G., & Deary, I. J. (2006). Effect of maternal smoking during pregnancy on offspring's cognitive ability: Empirical evidence for complete confounding in the US National Longitudinal Survey of Youth. *Pediatrics*, *118*, 943–950.

Brennan, P. A., Grekin, E. R., & Mednick, S. A. (1999). Maternal smoking during pregnancy and adult male criminal outcomes. *Archives of General Psychiatry*, *56*, 215–219.

Butler, N., & Goldstein, H. (1973). Smoking in pregnancy an subsequent child development. *British Medical Journal*, *4*, 573–575.

Cnattingius, S. (2004). The epidemiology of smoking during pregnancy: Smoking prevalence, maternal characteristics, and pregnancy outcomes. *Nicotine and Tobacco Research*, *6*, S125–S140.

Coe, C. L., & Lubach, G. R. (2008). Fetal programming: Prenatal origins of health and illness. *Current Directions in Psychological Science*, *17*, 36–41.

Cornelius, M. D., Ryan, C. M., Day, N. L., Godschmidt, L., & Willford, J. A. (2001). Prenatal tobacco effects on neuropsychological outcomes among preadolescents. *Journal of Developmental and Behavioral Pediatrics*, *22*, 217–225.

D'Onofrio, B. M., Turkheimer, E., Eaves, L. J., Corey, L. A., Berg, K., Solaas, M. H., et al. (2003). The role of the Children of Twins design in elucidating causal relations between parent characteristics and child outcomes. *Journal of Child Psychology and Psychiatry*, *44*, 1130–1144.

D'Onofrio, B. M., Turkheimer, E. N., Emery, R. E., Slutske, W., Heath, A., Madden, P. A. F., et al. (2005). A genetically informed study of marital instability and its association with offspring psychopathology. *Journal of Abnormal Psychology*, *114*, 570–586.

D'Onofrio, B. M., Van Hulle, C. A., Waldman, I. D., Rodgers, J. L., Harden, K. P., Rathouz, P. J., et al. (2008). Smoking during pregnancy and offspring externalizing problems: An exploration of genetic and environmental confounds. *Development and Psychopathology*, *20*, 139–164.

D'Onofrio, B. M., Van Hulle, C. A., Waldman, I. D., Rodgers, J. L., Rathouz, P. J., & Lahey, B. B. (2007). Causal inferences regarding prenatal alcohol exposure and childhood externalizing problems. *Archives of General Psychiatry*, *64*, 1296–1304.

Fergusson, D. M. (1999). Prenatal smoking and antisocial behavior. *Archives of General Psychiatry*, *56*, 223–224.

Fergusson, D. M., & Lloyd, M. (1991). Smoking during pregnancy and its effects on child cognitive ability from the ages of 8 to 12 years. *Paediatric and Perinatal Epidemiology*, *5*, 189–200.

Fogelman, K., & Manor, O. (1988). Smoking in pregnancy and subsequent development into early adulthood. *British Medical Journal*, *297*, 1233–1236.

Greene, W. H. (2003). *Econometric analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Harden, K. P., Lynch, S. K., Turkheimer, E., Emery, R. E., D'Onofrio, B. M., Slutske, W. S., et al. (2007). A behavior genetic investigation of adolescent motherhood and offspring mental health problems. *Journal of Abnormal Psychology*, *116*, 667–683.

Heath, A. C., Kendler, K. S., Eaves, L. J., & Markell, D. (1985). The resolution of cultural and biological inheritance: Informativeness of different relationships. *Behavior Genetics*, *15*, 439–465.

Huizink, A. C., & Mulder, E. J. H. (2006). Maternal smoking, drinking or cannabis use during pregnancy and neurobehavioral and cognitive functioning in human offspring. *Neuroscience and Biobehavioral Reviews*, *30*, 24–41.

Knopik, V. S., Sparrow, E. P., Madden, P. A. F., Bucholz, K. K., Hudziak, J. J., Reich, W., et al. (2005). Contributions of parental alcoholism, prenatal substance exposure, and genetic transmission to child ADHD risk: A female twin study. *Psychological Medicine*, *35*, 625–635.

Krueger, R. F., Moffitt, T. E., Caspi, A., Bleske, A., & Silva, P. A. (1998). Assortative mating for antisocial behavior: Developmental and methodological implications. *Behavior Genetics*, *28*, 173–186.

Lahey, B. B., D'Onofrio, B. M., & Waldman, I. D. (2009). Using epidemiologic methods to test hypotheses regarding causal influences on child and adolescent mental disorders. *Journal of Child Psychology and Psychiatry*, *50*, 53–62.

Lambe, M., Hultman, C., Torrang, A., MacCabe, J., & Cnattingius, S. (2006). Maternal smoking during pregnancy and school performance at age 15. *Epidemiology*, *17*, 524–530.

Lichtenstein, P., Pedersen, N. L., & McClearn, G. E. (1992). The origins of individual differences in occupational status and educational level. *Acta Sociologica*, *35*, 13–31.

Lindmark, G., & Cnattingius, S. (1991). The scientific basis of antenatal care. Report from a state-of-the-art conference. *Acta Obstetricia et Gynecologica Scandinavica*, *70*, 105–109.

Maughan, B., Taylor, A., Caspi, A., & Moffitt, T. E. (2004). Prenatal smoking and early childhood conduct problems. *Archives of General Psychiatry*, *61*, 836–843.

Moffitt, T. E. (2005). The new look of behavioral genetics in developmental psychopathology: Gene-environment interplay in antisocial behaviors. *Psychological Bulletin*, *131*, 533–554.

Muthén, L. K., & Muthén, B. O. (1998–2007). *Mplus user's guide* (5th ed.). Los Angeles: Author.

Neuhaus, J. M., & McCulloch, C. E. (2006). Separating between- and within-cluster covariate effects by using conditional and partitioning methods. *Journal of the Royal Statistical Society, 68*, 859–872.

Rantakallio, P. (1983). A follow-up study up to the age of 14 of children whose mothers smoked during pregnancy. *Acta Paediatrica Scandinavica, 72*, 747–753.

Reiss, D., Neiderhiser, J. M., Hetherington, E. M., & Plomin, R. (2000). *The relationship code: Deciphering genetic and social patterns in adolescent development*. Cambridge, MA: Harvard University Press.

Rodgers, J. L., Cleveland, H., van den Oord, E., & Rowe, D. (2000). Resolving the debate over birth order, family size, and intelligence. *American Psychologist, 55*, 599–612.

Rutter, M. (2007). Proceeding from observed correlation to causal inference: The use of natural experiments. *Perspectives on Psychological Science, 2*, 377–395.

Rutter, M., Pickles, A., Murray, R., & Eaves, L. J. (2001). Testing hypotheses on specific environmental causal effects on behavior. *Psychological Bulletin, 127*, 291–324.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147–177.

Silberg, J. L., & Eaves, L. J. (2004). Analyzing the contribution of genes and parent-child interaction to childhood behavioral and emotional problems: A model for the children of twins. *Psychological Medicine, 34*, 347–356.

Silberg, J. L., Parr, T., Neale, M. C., Rutter, M., Angold, A., & Eaves, L. J. (2003). Maternal smoking during pregnancy and risk to boys' conduct disturbance: An examination of the causal hypothesis. *Biological Psychiatry, 53*, 130–135.

Statistics Sweden. (1982). *Swedish socioeconomic classification: Reports on statistical coordination*. Orebro, Sweden: Statistics Sweden.

Statistics Sweden. (2003). *Multi-Generation Register 2002: A description of contents and quality*. Örebro, Sweden: Statistics Sweden.

Swedish Centre for Epidemiology. (2001). *The Swedish Medical Birth Registry: A summary of content and quality*. Retrieved November 15, 2009, from http://www.sos.se/fulltext/112/2003-112-3/summary.htm

Swedish National Agency for Education (n.d.). Retrieved November 15, 2009, from http://www.skolverket.se/

van den Oord, E. J., Simonoff, E., Eaves, L. J., Pickles, A., Silberg, J., & Maes, H. (2000). An evaluation of different approaches for behavior genetic analyses with psychiatric symptom scores. *Behavior Genetics, 30*, 1–18.

Wakschlag, L. S., Pickett, K. E., Cook, E., Benowitz, N. L., & Leventhal, B. L. (2002). Maternal smoking during pregnancy and severe antisocial behavior in offspring: A review. *American Journal of Public Health, 92*, 966–974.

Appendix

Table A1

Mean Offspring Outcomes Based on Exposure to Maternal Smoking During Pregnancy and Genetic Relatedness to Comparison Group (Including Offspring of Concordant Aunts and Siblings Concordant for Exposure to SDP)

	Grades		Math scores	
SDP	*M*	*n*	*M*	*n*
Unrelated offspring				
Unexposed	0.19	210,858	0.14	97,286
Exposed	−0.30	79,954	−0.24	31,417
Cousin comparison (all cousins)				
Concordant—no history SDP	0.20	90,167	0.17	40,357
Discordant—no history SDP	−0.07	30,656	−0.04	13,618
Discordant—history of SDP	−0.32	31,273	−0.24	12,440
Concordant—history of SDP	−0.48	29,003	−0.36	11,207
Cousin comparison (aunts were full siblings, r_g = .125)				
Concordant—no history SDP	0.22	85,985	0.18	38,550
Discordant—no history SDP	−0.04	26,524	−0.02	11,839
Discordant—history of SDP	−0.28	26,794	−0.22	10,667
Concordant—history of SDP	−0.45	25,280	−0.34	9,771
Cousin comparison (aunts were half siblings, r_g = .0625)				
Concordant—no history SDP	−0.10	4,182	−0.08	1,807
Discordant—no history SDP	−0.23	3,313	−0.14	1,428
Discordant—history of SDP	−0.55	3,351	−0.41	1,322
Concordant—history of SDP	−0.64	3,723	−0.47	1,436

Table A1
Continued

SDP	Grades		Math scores	
	M	*n*	*M*	*n*
Sibling comparison (all siblings)				
Concordant—unexposed	0.19	225,645	0.14	101,881
Discordant—unexposed	−0.19	16,467	−0.16	8,117
Discordant—exposed	−0.16	16,465	−0.18	5,423
Concordant—exposed	0.41	68,584	−0.33	27,138
Sibling comparison (full siblings, r_g = .50)				
Concordant—unexposed	−0.20	219,716	0.15	99,354
Discordant—unexposed	−0.16	14,675	−0.15	7,356
Discordant—exposed	−0.12	14,579	−0.14	4,801
Concordant—exposed	−0.37	61,041	−0.30	24,262
Sibling comparison (half siblings, r_g = .25)				
Concordant—unexposed	−0.36	5,929	−0.26	2,527
Discordant—unexposed	−0.54	1,377	−0.34	583
Discordant—exposed	−0.63	1,396	−0.51	525
Concordant—exposed	−0.76	7,543	−0.56	2,876

Note. SDP = smoking during pregnancy.

Table A2
Full Results for Model 2 for Average Grades

Risk factor	*b*	*SE*	*p*
SDP	−.201	0.004	.000
Maternal age at childbirth (years)	.007	0.000	.000
Paternal age at childbirth (years)	.000	0.000	.693
Gestational age (days)	.000	0.000	.010
Year of birth (years)	.021	0.001	.000
Not cohabiting	.158	0.012	.000
Offspring sex	.393	0.003	.000
Maternal less than 9 years of education	—	—	—
9 years of education	.048	0.015	.002
1–2 years upper secondary education	.206	0.015	.000
3 years upper secondary education	.345	0.015	.000
Less than 3 years postsecondary education	.440	0.015	.000
3+ years postsecondary education	.529	0.015	.000
Postgraduate education	.663	0.027	.000
Maternal unskilled blue collar	—	—	—
Skilled blue collar	.109	0.005	.000
Low-level white collar	.197	0.005	.000
Intermediate-level white collar	.200	0.006	.000
High-level white collar	.289	0.008	.000
Self-employed	.188	0.011	.000
Employed, uncategorized	.047	0.007	.000
Not working/missing	.031	0.008	.000
Maternal history of criminal conviction	−.176	0.005	.000
Paternal less than 9 years of education			
9 years of education	.026	0.009	.003
1–2 years upper secondary education	.100	0.008	.000
3 years upper secondary education	.266	0.009	.000
Less than 3 years postsecondary education	.325	0.009	.000
3+ years postsecondary education	.440	0.010	.000

Table A2
Continued

Risk factor	*b*	*SE*	*p*
Postgraduate education	.593	0.016	.000
Paternal unskilled blue collar			
Skilled blue collar	.044	0.005	.000
Low-level white collar	.116	0.006	.000
Intermediate-level white collar	.146	0.006	.000
High-level white collar	.194	0.007	.000
Self-employed	.137	0.007	.000
Employed, uncategorized	.011	0.008	.191
Not working/missing	.046	0.010	.000
Paternal history of criminal conviction	−.164	0.003	.000

Note. SDP = smoking during pregnancy.

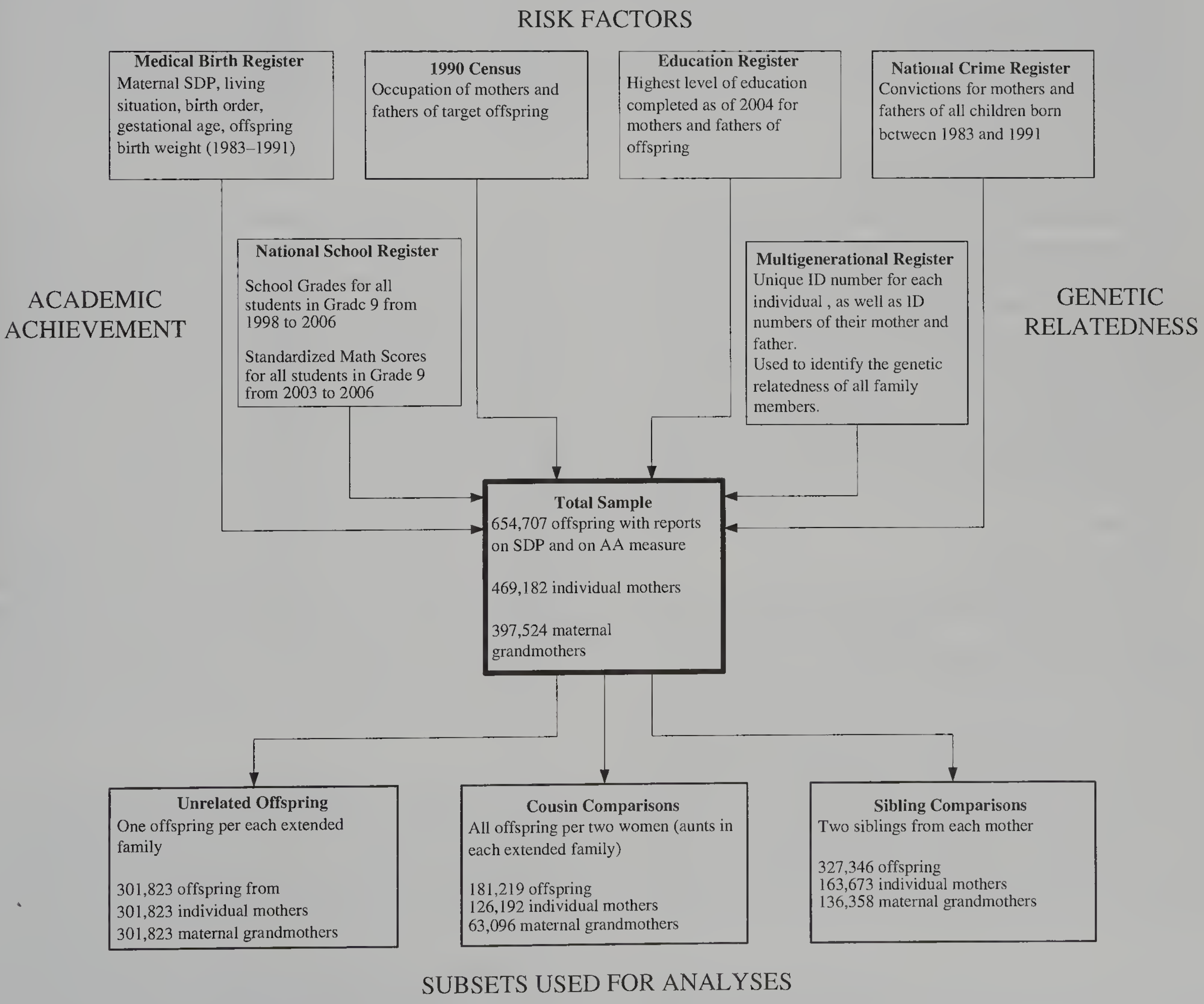

RISK FACTORS
Medical Birth Register
Maternal SDP, living situation, birth order, gestational age, offspring birth weight (1983–1991)
1990 Census
Occupation of mothers and fathers of target offspring
Education Register
Highest level of education completed as of 2004 for mothers and fathers of offspring
National Crime Register
Convictions for mothers and fathers of all children born between 1983 and 1991
ACADEMIC ACHIEVEMENT
National School Register
School Grades for all students in Grade 9 from 1998 to 2006
Standardized Math Scores for all students in Grade 9 from 2003 to 2006
Multigenerational Register
Unique ID number for each individual , as well as ID numbers of their mother and father.
Used to identify the genetic relatedness of all family members.
GENETIC RELATEDNESS
Total Sample
654,707 offspring with reports on SDP and on AA measure
469,182 individual mothers
397,524 maternal grandmothers
Unrelated Offspring
One offspring per each extended family
301,823 offspring from
301,823 individual mothers
301,823 maternal grandmothers
Cousin Comparisons
All offspring per two women (aunts in each extended family)
181,219 offspring
126,192 individual mothers
63,096 maternal grandmothers
Sibling Comparisons
Two siblings from each mother
327,346 offspring
163,673 individual mothers
136,358 maternal grandmothers
SUBSETS USED FOR ANALYSES

Figure A1. Explanation of linkage of national registries and creation of subsets for analyses.

Table A3

Full Results for Model 2 for Mathematics Proficiency

Risk factor	*b*	*SE*	*p*
SDP	−.148	0.006	.000
Maternal age at childbirth (years)	.012	0.001	.000
Paternal age at childbirth (years)	.001	0.001	.200
Gestational age (days)	.001	0.000	.000
Year of birth (years)	−.035	0.002	.000
Not cohabiting	.100	0.018	.000
Offspring sex	.034	0.005	.000
Maternal less than 9 years of education	—	—	—
9 years of education	−.019	0.031	.536
1–2 years upper secondary education	.080	0.030	.007
3 years upper secondary education	.192	0.030	.000
Less than 3 years postsecondary education	.296	0.030	.000
3+ years postsecondary education	.395	0.031	.000
Postgraduate education	.498	0.045	.000
Maternal unskilled blue collar	—	—	—
Skilled blue collar	.061	0.008	.000
Low-level white collar	.127	0.008	.000
Intermediate-level white collar	.135	0.008	.000
High-level white collar	.226	0.012	.000
Self-employed	.159	0.017	.000
Employed, uncategorized	.046	0.010	.000
Not working/missing	.005	0.011	.611
Maternal history of criminal conviction	−.106	0.008	.000
Paternal less than 9 years of education			
9 years of education	.038	0.015	.011
1–2 years upper secondary education	.095	0.014	.000
3 years upper secondary education	.248	0.015	.000
Less than 3 years postsecondary education	.323	0.016	.000
3+ years postsecondary education	.435	0.016	.000
Postgraduate education	.597	0.026	.000
Paternal unskilled blue collar	—	—	—
Skilled blue collar	.027	0.007	.000
Low-level white collar	.086	0.009	.000
Intermediate-level white collar	.131	0.008	.000
High-level white collar	.183	0.010	.000
Self-employed	.129	0.011	.000
Employed, uncategorized	.038	0.012	.001
Not working/missing	−.010	0.014	.505
Paternal history of criminal conviction	−.096	0.005	.000

Note. SDP = smoking during pregnancy.

Child Development, January/February 2010, Volume 81, Number 1, Pages 101–114

Fetal Genotype for the Xenobiotic Metabolizing Enzyme *NQO1* Influences Intrauterine Growth Among Infants Whose Mothers Smoked During Pregnancy

Thomas S. Price
Kings College London

Tilo Grosser
University of Pennsylvania

Robert Plomin and Sara R. Jaffee
Kings College London

Maternal smoking during pregnancy retards fetal growth and depresses infant birth weight. The magnitude of these effects may be moderated by fetal genotype. The current study investigated maternal smoking, fetal genotype, and fetal growth in a large population sample of dizygotic twins. Maternal smoking retarded fetal growth in a dose-dependent fashion. In a subsample of 497 twin pairs whose mothers smoked during pregnancy, a functional polymorphism in the NAD(P)H:quinone oxidoreductase gene (*NQO1* Pro187Ser; rs1800566) was significantly associated with fetal growth within families. The effect was strongest among moderate smokers. This is the first demonstration that fetal genotype for an enzyme involved in tobacco smoke metabolism influences intrauterine growth independent of maternal genotype. Future studies should conduct formal tests of Fetal Genotype × Maternal Smoking interactions.

Maternal smoking during pregnancy is a well-established and preventable risk factor for low birth weight (< 2,500 g) and its sequelae of poor physical, cognitive and behavioral development, excess morbidity, and increased rates of both perinatal and adult mortality (Kramer, 2003). The primary causes of low birth weight are preterm birth and intrauterine growth restriction (IUGR). The consequences of IUGR include both short-term and long-term morbidity and permanent deficits in growth and neurocognitive development (Kramer, 2003). Epidemiological studies have shown that maternal smoking is associated with both short gestation and IUGR (Kramer, 2003). Mothers who quit smoking while pregnant have longer gestations and heavier newborns than those who continue to smoke (Lumley, Oliver, Chamberlain, & Oakley, 1998). This fact is recognized in public policy: In the United Kingdom, formal smoking cessation programs are recommended as part of antenatal care to prevent low birth weight (Health Development Agency, 2004). This study investigated the effects of fetal genotype on intrauterine growth (IUG) for infants whose mother smoked during pregnancy. The goal was to prioritize candidates for future epidemiological and functional investigation of fetal genotypes that modify the effects of tobacco smoke exposure on fetal growth.

The Role of Xenobiotic Metabolizing Enzymes in Prenatal Smoke Exposure

The placenta provides nutrients to the fetus and acts as a barrier to protect the fetus from foreign substances (known as xenobiotics) in the maternal blood. Direct toxicity in developing fetal cells and the depression of placental function by smoke components are critical factors in the retardation of fetal growth. Mechanistically, smoke components can cause oxidative chemical stress in placental and fetal cells, reduce the placental transport of blood

---

The authors thank Kathakali Addya for supervising the genotyping, Madhavi Vaddi for performing the genotyping assays, Charles Curtis for a great deal of technical laboratory assistance, Andy McMillan for data management, and Steve Whitehead for inspiration and encouragement. This work was supported by grants from the Merck Foundation (to S.R.J.) and from the American Heart Association (National Scientist Development Grant 0430148N to T.G.).

Correspondence concerning this article should be addressed to Thomas S. Price, P080 SGDP Centre, Institute of Psychiatry, Kings College London, London SE5 8AF, UK. Electronic mail may be sent to thomas.price@kcl.ac.uk..

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8101-0007

or nutrients, perturb endocrine placental functions, and induce fetal hypoxemia (blood oxygen deficiency). The authors hypothesize that chemical stress on fetal cells may provide a tractable target in terms of identifying fetal genetic modifiers of tobacco smoke exposure.

Tobacco smoke is a complex mixture of compounds including three important classes of procarcinogens: polycyclic aromatic hydrocarbons, arylamines, and *N*-nitrosamines (Bartsch et al., 2000). Both maternal and fetal cells, in the placenta and other tissues, defend against these harmful substances with a biochemical one-two punch: enzyme-catalyzed Phase I (activation) and Phase II (detoxification) reactions. Phase I enzymes typically introduce a reactive group into the molecule that enables its conjugation in a Phase II reaction. The Phase I intermediate is often more harmful than its precursor and may promote oxidant stress or damage DNA molecules, processes that may mediate the effects of tobacco smoke on fetal growth (Kim et al., 2005; Li & Wang, 2004; Myatt & Cui, 2004). Thus, it is likely that IUG under conditions of smoke exposure is critically dependent on the efficacy with which maternal, fetal, and placental cells can metabolize and eliminate exposure toxins. Indeed, polymorphisms in the xenobiotic metabolizing enzymes that activate and detoxify the components of tobacco smoke have been shown to influence its effects on fetal development. Associations have been reported between fetal growth and maternal variants in genes including *CYP1A1* (Chen et al., 2005; Infante-Rivard, Weinberg, & Guiguet, 2006; Sasaki et al., 2006; Sram et al., 2006; Wang et al., 2002; Wu et al., 2007), *GSTT1* (Infante-Rivard et al., 2006; Wang et al., 2002), *GSTM1* (Sasaki et al., 2006), *AhR* (Sasaki et al., 2006), *CYP2E1* (Sasaki et al., 2008), and *NQO1* (Sasaki et al., 2008). While one would expect that fetal genotype would also modify the effects of maternal smoking, this has only begun to be investigated. Thus, a recent study that assessed the interaction between maternal smoking and fetal genotype reported that newborns carrying one or two deletions of *GSTT1* were at increased risk of being small for gestational age (SGA) if their mothers smoked during pregnancy (Infante-Rivard et al., 2006). These results, however, were not adjusted for the correlated effects of maternal genotype, so it remains unknown whether fetal genotype for xenobiotic metabolizing enzyme genes makes an independent contribution to the susceptibility to smoking-related growth retardation.

The Role of Folate in Prenatal Smoke Exposure

Availability of folate to the fetus in early gestation may also be a significant determinant of infant birth weight and is influenced by both genetic factors and environmental conditions, including smoking. Folate is a micronutrient with critical importance for fetal development. Reduced maternal folate status during pregnancy strongly increases the risk of neural tube defects (Lumley, Watson, Watson, & Bower, 1998). Observational studies (Mitchell et al., 2004; Neggers, Goldenberg, Tamura, Cliver, & Hoffman, 1997; Rao et al., 2001) suggest that folate supplementation may reduce the risk of IUGR and low infant birth weight. Data from controlled trials suggest that folic acid at high doses reduces the risk of low birth weight (pooled relative risk at 0.73; 95% confidence interval [CI] 0.53, 0.99), suggesting that the association is causal (Charles et al., 2005). Studies that have measured maternal folate status directly have also found associations with IUGR (Goldenberg et al., 1992; Steegers-Theunissen, Van Iersel, Peer, Nelen, & Steegers, 2004) and infant birth weight (Relton, Pearce, & Parker, 2005). Moreover, cigarette smokers have lower levels of folate in whole blood, red blood cells, plasma, and serum (Kato et al., 1999; Refsum et al., 2006; Walmsley, Bates, Prentice, & Cole, 1999). Maternal smoking also depletes infant folate (Stark, Pawlosky, Sokol, Hannigan, & Salem, 2007). It is likely that smoking aggravates the effects of poor folate status on fetal growth: There is evidence that the relation between plasma folate levels and IUGR may be stronger in mothers who smoke (Sram, Binkova, Lnenickova, Solansky, & Dejmek, 2005).

The metabolism of folate in vivo is known to be influenced by genetic variation (Fredriksen et al., 2007). The link between folate-related genetic variation and fetal growth, however, is far from clear (Engel, Olshan, Siega-Riz, Savitz, & Chanock, 2006; Franchi et al., 2004; Gebhardt, Scholtz, Hillermann, & Odendaal, 2001; Glanville et al., 2003; Glanville et al., 2006; Infante-Rivard et al., 2002; Kupferminc et al., 1999; Martinelli et al., 2001; McCowan et al., 2003; Murphy et al., 2000; Nurk, Tell, Refsum, Ueland, & Vollset, 2004; Park et al., 2005; Relton, Pearce, Burn, & Parker, 2005; Salomon et al., 2004; Verspyck et al., 2002). The general lack of significant main effects of folate pathway genes on IUGR, however, does not rule out the possibility that folate-related polymorphic variation influences fetal growth by moderating the effects of environmental risk factors such as low folate intake. Several studies have reported associations between folate pathway genotype and fetal growth only among mothers with depleted folate status (Engel et al., 2006; Infante-Rivard et al., 2002; Relton, Pearce, Burn, et al., 2005; one study has reported a significant genotype–environment interaction with maternal folate status, Park et al., 2005). In addition, a fetal polymorphism in a folate pathway gene has been shown to moderate the effects of maternal smoking on risk for neural tube defects (Jensen, Hoess, Whitehead, & Mitchell, 2005). Although the causal mechanisms remain unknown, it is plausible that poor neurodevelopmental outcomes arise because folate genotype and maternal smoking interact mechanistically to reduce levels of circulating folate. Indeed, it has been reported that *MTHFR* C677T, a functional polymorphism in the folate pathway gene methylenetetrahydrofolate reductase, moderates the effect of tobacco exposure on serum folate levels (McDonald, Perkins, Jodouin, & Walker, 2002).

A powerful method of prioritizing genetic candidates when a Gene × Environment interaction (G × E) is suspected is to search for genetic association in a sample of subjects who have been exposed to the relevant environmental risk (Moffitt, Caspi, & Rutter, 2005). The current study tested fetal genetic variants in xenobiotic metabolizing enzymes and folate pathway genes for association with fetal growth in a population sample of dizygotic (DZ) twins born to mothers who smoked during their pregnancy. The focus was on common polymorphisms with known or suspected functional activity in genes responsible for metabolizing components of tobacco smoke (see Table 1). These included the Phase I enzymes epoxide hydrolase (*EPHX1*), cytochrome P450 isoforms *CYP1A2* and *CYP1B1*, the Phase II enzyme glutathione *S*-transferase P (*GSTP1*), and NAD(P)H quinone oxidoreductase (*NQO1*), an enzyme with both Phase I and Phase II activity. It was also hypothesized that the effect of maternal smoking to depress birth weight was moderated by polymorphisms in genes involved in the metabolism of folate, including *MTHFR*, methionine synthase (*MTR*), methionine synthase reductase (*MTRR*), methylenetetrahydrofolate dehydrogenase (*MTHFD1*), cystathione β-synthase (*CBS*), and cytosolic serine hydroxymethyltransferase (*SHMT1*; see Table 1). In addition, it was anticipated that the strength of any genetic association might vary with the dose of tobacco exposure.

Comparing outcomes within families ensured matching within twin pairs for their exposure to tobacco smoke and other risk factors that might account for the association between smoking during pregnancy and IUG. Consequently, this research design eliminated many of the potential confounds present in previous studies. Although

Table 1

Allele Frequencies, Tests of Hardy–Weinberg Equilibrium (HWE), and Tests of Association Between Genotype and Intrauterine Growth

Gene	SNP	Major/minor allele	MAF	HWE (*p* value)	Heterogeneity −2*LL*	Total −2*LL*	Within families −2*LL*
Xenobiotic metabolizing enzyme genes							
CYP1B1	rs1056836	G/C	.43	.50	1.58	0.54	0.44
CYP1B1	rs1800440	T/C	.19	.45	2.22	3.01	4.32
CYP1A2	rs2470890	A/G	.36	.14	0.57	2.60	3.04
NQO1	rs1800566	C/T	.19	.15	6.88*	—	11.76**
GSTP1	rs1695	A/G	.34	.61	2.81	2.06	2.05
GSTP1	rs1138272	C/T	.09	.30	9.37*	—	3.59
EPHX1	rs1051740	A/G	.31	.91	4.08	4.19	0.80
EPHX1	rs2234922	A/G	.19	.06	1.00	0.11	0.69
Folate pathway genes							
MTHFR	rs1801133	G/A	.32	.23	0.52	3.40	1.15
MTHFR	rs1801131	A/C	.31	.62	5.20	0.77	4.71
MTR	rs1805087	A/G	.20	.43	3.27	0.50	1.24
MTRR	rs1801394	C/T	.44	.80	0.00	2.99	0.00
MTHFD1	rs2236225	G/A	.48	.34	2.25	0.57	0.03
CBS	rs234706	G/A	.36	.65	5.39	0.12	3.12
SHMT1	rs1979277	G/A	.31	.11	2.92	1.39	3.42

Note. SNP = single nucleotide polymorphism; MAF = minor allele frequency.
$^{*}p < .05$. $^{**}p < .0033$ ($p < .05$ after Bonferroni correction for 15 tests).

twins tend to have lower birth weights and shorter gestations than singletons due to crowding in the uterus, prenatal exposure to tobacco is reliably associated with lower birth weight (Pollack, Lantz, & Frohna, 2000; Salihu, Aliyu, & Kirby, 2005), and low birth weight is associated with elevated risk of poor developmental outcomes (Aylward, 2005; Rooney, Hay, & Levy, 2003) in both twins and singletons. Because the current study defined IUG as birth weight adjusted for gestational age, the findings may be more likely to generalize to singletons (Morley, 2005).

Method

Sample

The sample is a subset of the Twins Early Development Study (TEDS), a study of all twins born in England and Wales in 1994, 1995, and 1996 (Trouton, Spinath, & Plomin, 2002). From 25,815 live twin births identified by the government organization responsible for birth records, 65.5% of families were contacted, and of these, 81% agreed to participate. Each family received a questionnaire booklet around 18 months after the birth of the twins asking for details about the pregnancy and birth, the health of family members, structure of the family, and other demographic variables. In total, 13,429 families returned the booklet. The families who responded were reasonably representative of parents with young children in the UK population in terms of ethnicity, education, and employment status (Price, 2003). Of these, 6,942 families provided DNA samples for the twins, collected from buccal swabs (Freeman et al., 1997). Zygosity was determined by examining the identity of polymorphic DNA markers. Where DNA was unavailable, zygosity was determined using a standard parental report questionnaire asking about the physical similarity of the twins that has been shown to have 95% accuracy (Price et al., 2000). The full sample ($n = 8{,}918$ pairs) included in these analyses was restricted to DZ pairs and contained 2,233 male–male DZ pairs, 2,123 female–female DZ pairs, and 4,562 male–female pairs.

A subsample of twins whose mothers smoked during pregnancy was selected for genotyping. DNA was available from a total of 145 DZ male–male pairs, 153 DZ female–female pairs, and 246 male–female pairs whose mothers reported smoking during pregnancy in the booklet and who provided DNA samples. One twin pair was excluded from the analysis due to invalid birth weight data. A further 6 pairs were excluded whose gestational age was unknown, and 10 pairs were excluded because maternal alcohol consumption was not reported. A further 30 pairs were excluded because a majority of the genotyping assays failed in the first pass of genotyping for one or other twin, leaving a final sample of 497 twin pairs.

Measures

Birth weight and IUG. Birth weight was reported by parents in the booklet and recorded in either pounds and ounces or in kilograms. These reports were converted to grams ($M = 2{,}516$ g, $SD = 549$ g). Reported birth weights of less than 500 g ($n = 11$) were excluded from the data: The World Health Organization recommends excluding children below this threshold from statistics for live births due to near 100% mortality (Cartlidge & Stewart, 1995). Reported birth weights in excess of 5,000 g ($n = 6$) were treated as extreme outliers and excluded from the data set. Male twins tended to be about 100 g heavier than female twins (DZ male–male pairs: $M = 2{,}549$ g, $SD = 551$ g; DZ female–female pairs: $M = 2{,}456$ g, $SD = 542$ g; males in opposite-sex pairs: $M = 2{,}581$ g, $SD = 562$ g; females in opposite-sex pairs: $M = 2{,}457$ g, $SD = 546$ g). IUG was defined as birth weight corrected for gestational age. The expected birth weight given gestational age was calculated using a nonparametric smoothing method (local quadratic regression with a smoothing span of 0.75). IUG ranged from −2,227 to 2,088 g ($M = -11$ g, $SD = 391$ g) and was approximately normally distributed after correcting for the effects of twin sex and birth order (skewness = 0.08, kurtosis = 1.03). Importantly, intrapair differences in IUG were also approximately normally distributed after correcting for sex and birth order (skewness = −0.01, kurtosis = 0.58).

Smoking during pregnancy. A section of the booklet sent to parents approximately 18 months after the twins were born was devoted to features of the mother's pregnancy with the twins and the twins' early life. One question asked the mother to report whether she smoked during the pregnancy; 19.1% of mothers responded that they did. If the mother endorsed this item, she was asked to report the trimester or trimesters in which she smoked and how many cigarettes per day she smoked on average. The number of cigarettes mothers reported smoking ranged from 1 to 40 ($M = 10.5$, $SD = 6.1$). For the purposes of these analyses, the quantity of tobacco was recoded into an ordinal variable with the following categories: smoked less than 10 cigarettes per day ("moderate smokers"; 36.9%), smoked 10 or more cigarettes per day ("heavy smokers"; 63.1%).

Demographic and maternal covariates. Several covariates were investigated as possible confounders of the relation between smoking and IUG. The covariates were dichotomous and measured at the family level. Frequencies of the covariates were: alcohol consumption (mother reported drinking 1 or more units of alcohol per week): 27.0%; ethnicity reported as other than "White": 7.1%; mother reported education to "A" level standard (a qualification equivalent to a high school diploma): 33.0%; twins were the oldest children in the family home (a proxy for primiparity): 47.7%; twins were delivered by Caesarian section: 46.2%; mother reported suffering from high blood pressure during the pregnancy: 24.4%; mother reported suffering from diabetes during the pregnancy: 2.6%; mother reported suffering from pre-eclampsia during the pregnancy (a condition that can be dangerous to both mother and baby, characterized by high blood pressure and the presence of protein in the urine): 9.5%.

Genotyping was performed on a subsample of the DZ offspring of mothers who smoked during pregnancy and for whom DNA samples were available, as described previously. Only common single nucleotide polymorphisms (SNPs) with proven or suspected functional activity in xenobiotic metabolizing enzymes and folate pathway genes were selected for genotyping. With one exception (rs234706) these were nonsynonymous variants causing an alteration in the amino acid sequence of the encoded protein. Non-SNP variants, such as insertions, deletions, and variable number repeat polymorphisms, could not be genotyped using the technology that was employed. SNPs with an estimated minor allele frequency in European populations of less than .1 were not genotyped because of low statistical power to test hypotheses relating to rare variants.

Genotyping was performed at the University of Pennsylvania School of Medicine Molecular Genotyping and Diagnosis Core Facility using a standard protocol (Rennert et al., 2005). TaqMan 5′ nuclease PCR primers and probes for alleles of the variant of interest were obtained from Applied Biosystems (ABI, Foster City, CA). Each probe consisted of an oligonucleotide with a fluorescent reporter dye, a nonfluorescent quencher, and minor groove binder (MGB). Allele-specific cleavage of probes was detected using different reporter dyes for each probe (6FAM and VIC fluorophores for each allele) with separate wavelength maxima. PCR amplifications were set up in a 384-well plate format in total volume of 5 μl, containing 50–900 nM of each primer, 50–250 nM of each probe, TaqMan Universal Master Mix (ABI) and constant dilutions of 5 ng of DNA sample for assays by design. PCR amplifications were set up in a 384-well plate format in total volume of 5 μl, containing 2.5 μl 2× universal master mix, 0.25 μl 1× primer and probe from ABI and 2.25 μl of DNA at a concentration of 5 ng/μl. Controls representing each genotype for each variant and a no template (water) control were included in each 384-well plate. PCR was performed in ABI 7900 HT Sequence Detection System (ABI). After an enzyme activation step for 10 min at 95°C, 60 two-step cycles were performed; 15 s denaturation at 95°C followed by 1 min annealing/extension at 60°C for all variants. After PCR endpoint, fluorescence levels of 6FAM and VIC were measured automatically in each well using the SDS 2.1 manufacturer's custom software (ABI). Allelic discrimination results were then graphed on a scatter plot contrasting reporter dye florescence for one allele against the other.

In 30 of the genotyped individuals (all in different families) a majority of the genotyping assays failed in the first pass of genotyping, probably because the DNA sample was low in volume or of poor quality. After excluding these families, no individual in the final sample had more than three genotyping failures. All of these failed assays were repeated twice. After repeating the failed genotyping assays, there were only three missing genotypes in the data set.

Statistical analysis proceeded in two parts. In the first part, the association between smoking during pregnancy and offspring IUG was tested in the overall sample. In the second part, tests of association between fetal genotype and IUG were conducted in the genotyped subsample using an adaptation of the statistical model implemented in the program QTDT (Abecasis, Cardon, & Cookson, 2000; Fulker, Cherny, Sham, & Hewitt, 1999). The QTDT model tests the association between IUG and genotype both within and between families assuming that IUG in twin pairs follows a bivariate normal distribution. The test of association between families is equivalent to calculating the average IUG values for each twin pair in the sample and regressing this on the average genotype values. The test of association within families is equivalent to regressing the within-pair differences in IUG on the within-pair differences in genotype. In this model, the genetic association between families is biased in the presence of stratification (confounding due to admixture between ethnic populations with differing allele frequencies and mean IUG), but the association within families is not. Furthermore, the mean sibling genotype, as estimated in tests of the between-families effect, is confounded with maternal genotype. As a result, the estimate of the genetic association with IUG between families may include effects of maternal genotype as well as fetal genotype. The QTDT model allows a test of the heterogeneity of between- and within-family effects (Abecasis et al., 2000; Fulker et al., 1999; implemented as the "stratification" association model in the original software, http://www.sph.umich.edu/csg/abecasis/QTDT): The null hypothesis, in this instance, is that the effects of stratification and maternal genotype are absent or cancel each other out. Under the null hypothesis, the effects of fetal genotype can be estimated both between and within families. Otherwise, only the within-family test is appropriate.

Birth weight was adjusted for gestational age, twin sex, and for a set of familial and gestational covariates measuring maternal alcohol consumption, ethnicity, maternal education, primiparity, birth by Caesarian section, maternal high blood pressure, diabetes, and toxemia/pre-eclampsia. Two coefficients were estimated for each of these covariates: one coefficient for its shared environmental effect and another coefficient for its effect on the birth order difference in birth weight. Similar coefficients were also estimated for the effects of smoking dose, considered as a binary variable (mother smoked 10 or more cigarettes per day). Likelihood ratio tests comparing nested models were performed to test the principal hypothesis of an additive main effect of genotype on IUG and an interactive effect of genotype with smoking dose (2 *df* tests). The 2 *df* test has been recommended as the most powerful method to screen for genetic associations when G × E is likely, but the true G × E model is not known (Kraft, Yen, Stram, Morrison, & Gauderman, 2007).

For each SNP, an initial test of the heterogeneity of between- and within-family effects was performed. If this test was nonsignificant using $\alpha = .05$, the effects of fetal genotype both within and between families were estimated according to the "total" model. If the test of heterogeneity was significant, the effects of fetal genotype were estimated within families only (the "within-families" model, equivalent to the "orthogonal" model in the original QTDT). Post hoc analyses tested the following hypotheses: (a) additive main effect of genotype (1 *df* test); (b) additive G × E in the presence of an additive main effect (1 *df* test); and (c) effects of genetic nonadditivity, following Posthuma, de Geus, Boomsma, and Neale (2004) (4 *df* test). All tests were conducted under the assumption of no linkage because there was insufficient information from the small number of SNPs tested to estimate the numbers of alleles identical by descent with accuracy.

Statistical analyses were conducted using the freely distributed software packages *R* for Windows version 2.6.1 (R Development Core Team, 2007) and *Mx* for Windows (*MxGUI* version 1.7.03; Neale, Boker, Xie, & Maes, 2003).

Results

Is Smoking During Pregnancy Associated With IUG in the TEDS Sample Overall?

In the sample overall, children whose mothers smoked weighed 152 g less at birth than children whose mothers did not smoke, $t(8{,}197\ df) = -16.7$, $p < .0001$, after controlling for gestational age and twin sex. The offspring of heavy smokers (10 or more cigarettes per day) weighed 167 g less than offspring of nonsmokers, whereas the offspring of moderate smokers weighed 126 g less than offspring of nonsmokers, after controlling for gestational age and twin sex, indicating a significant effect of tobacco dose on IUG, $t(8{,}196\ df) = -2.4$, $p = .01$.

Is the Relation Between Smoking During Pregnancy and IUG Confounded by Family or Maternal Characteristics in the TEDS Sample Overall?

After adjusting for the effects of twin sex, higher IUG was associated with white ethnicity, more educated mothers, primiparity, maternal diabetes, and the absence of toxemia and pre-eclampsia during the pregnancy (all $ps < .05$). Firstborn twins were, on average, 22 g heavier than their younger cotwin. The birth-order effect was greater for mothers who were not nulliparous (i.e., in families with older children), in mothers reporting high blood pressure and toxemia/pre-eclampsia, and in children born by Caesarian section, after adjusting for twin sex (all $p < .05$).

After controlling for the effects of twin sex and the entire set of covariates, the effect of maternal smoking on IUG remained statistically significant, reducing IUG by 159 g compared to the offspring of nonsmokers, $t(8{,}005\ df) = -17.3$, $p < .0001$. The additional impact of heavy smoking (a reduction of 185 g compared to nonsmokers) over moderate smoking (a reduction of 118 g compared to nonsmokers) also remained significant, $t(8{,}004\ df) = -3.9$, $p < .0001$.

Is Fetal Genotype Associated With IUG Among Twins Whose Mothers Smoked During Pregnancy?

Among mothers in the genotyped sample, 192 mothers (38.6%) reported smoking fewer than 10 cigarettes per day; 305 mothers (61.4%) reported smoking 10 cigarettes or more per day. These proportions were very similar to those for mothers in the entire sampling frame who reported smoking during the pregnancy. The dose effect of smoking was comparable to that in the sample overall. Heavy smoking was associated with a reduction in IUG of 62 g relative to moderate smoking, after controlling for gestational age, twin sex, and maternal and familial characteristics, $t(486\ df) = -1.9$, $p = .05$.

The genotype frequencies and test of Hardy–Weinberg Equilibrium (HWE) are shown in Table 1. Minor allele frequencies (MAF) ranged from .09 to .48. None of the SNPs deviated from HWE at $\alpha = .05$ according to the QLHW test that is valid in samples of related individuals (Bourgain, Abney, Schneider, Ober, & McPeek, 2004). The two markers in *MTHFR* were in linkage disequilibrium although they were not highly correlated ($D' = 1.00$, $r^2 = .22$); the same was true for the two SNPs in *CYP1B1* ($D' = .99$, $r^2 = .18$). In contrast, there was evidence of substantial recombination between the two polymorphisms in the *EPHX1* gene ($D' = 0.03$, $r^2 = .00$).

Tests of genetic association with IUG were conducted using a modification of the QTDT regression model (Posthuma et al., 2004) that allows for moderation of the genetic effect by smoking dose (see Table 1). There was evidence for a significant association between *NQO1* Pro187Ser (rs1800566) and IUG within families, $-2LL\ (2\ df) = 11.76$, $p = .0028$. This result remained statistically significant ($p = .042$) after Bonferroni correction for multiple testing (15 tests). The association between the variant *NQO1* allele and IUG within families reached nominal levels of statistical significance when the data were reanalyzed using reported tobacco intake as a quantitative variable, $-2LL\ (2\ df) = 8.92$, $p = .012$. Post hoc testing revealed that the main, $-2LL\ (1\ df) = 7.79$, $p = .0052$, and interactive effects, $-2LL\ (1\ df) = 3.96$, $p = .047$, of the rs1800566 SNP were both statistically significant. In twins whose mothers were moderate smokers during pregnancy (less than 10 cigarettes per day), possession of one more Ser allele than one's cotwin was associated with a 178 g elevation in birth weight relative to the cotwin, after correcting for twin sex, maternal and familial characteristics, and birth-order effects (see Figure 1). The association was absent in the offspring of heavy smokers. There was no evidence of genetic nonadditivity. Both the main effect and the interaction remained significant when the set of familial covariates was removed from the model. No other SNPs achieved even nominal levels of significance.

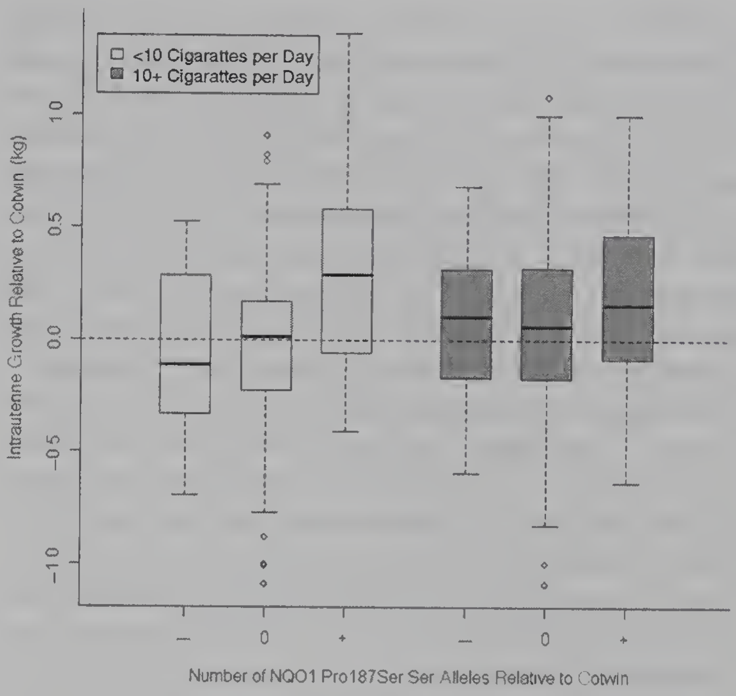


Figure 1. Effect of *NQO1* Pro187Ser (rs1800566) genotype on intrauterine growth of firstborn twin relative to cotwin for offspring of moderate smokers (less than 10 cigarettes per day; white boxes) and offspring of heavy smokers (10 or more cigarettes per day; gray boxes). Intrauterine growth is adjusted for gestational age and twin sex.

Discussion

Maternal smoking during pregnancy is known to retard fetal growth and increase the incidence of low birth weight. The relations between maternal smoking, fetal genotype, and fetal growth were investigated in a large population sample of DZ twins. Maternal smoking retarded growth by 118 g in twins born to mothers who reported smoking less than 10 cigarettes per day, allowing for the effects of twin sex and gestational age, and by 185 g in twins born to heavier smokers. A subsample of 497 twin pairs whose mothers smoked during pregnancy were selected for a molecular genetic study. In this subsample, a functional SNP in the *NQO1* gene (Pro187Ser; rs1800566) was significantly associated with fetal growth within families assuming a dose-related model of effects ($p = .0028$). Post hoc testing within families suggested that possession of the fetal Ser allele protected against fetal growth retardation in the offspring of moderate smokers but not heavy smokers (interaction $p = .047$). These results provide the first demonstration that fetal genotype for a xenobiotic metabolizing enzyme influences IUG under conditions of smoke exposure independently of maternal genotype. This finding implicates the importance to fetal development of Phase I activation of smoke xenobiotics by *NQO1* by cells in the growing fetus or fetal cells in the placenta.

The *NQO1* Pro187Ser variant results from a C to T substitution in exon 6 that codes for a proline to serine change at position 187 in the amino acid sequence. The mutant form of the protein has reduced enzymatic activity (Traver et al., 1997) and is rapidly degraded (Siegel et al., 2001). There is a gene–dose relation between the mutant allele and levels of *NQO1* protein in vivo (Siegel, McGuinness, Winski, & Ross, 1999) and *NQO1* activity in cancer cell lines (Ross & Siegel, 2004): In homozygotes the protein is virtually absent, whereas heterozygotes have intermediate levels of protein concentration and enzymatic activity.

NQO1 is a xenobiotic metabolizing enzyme that catalyzes both Phase I (activation) and Phase II (detoxification) reactions with certain chemicals in tobacco smoke and plays a role in stabilizing an enzyme (p53) that protects against cancer (Ross & Siegel, 2004). The Ser allele may protect against the effects of smoking during pregnancy because the inactive form of the *NQO1* protein produced by Ser heterozygotes and homozygotes does not activate toxic compounds causing DNA damage or oxidative stress (De Flora, Bennicelli, Dagostini, Izzotti, & Camoirano, 1994; Hajos & Winston, 1991; Knox et al., 1988). The protective effects of the Ser allele may be offset by decreased participation in phase II reactions, in which *NQO1* prevents oxidative stress by helping detoxify reactive intermediates (Cadenas, 1995).

The effects of *NQO1* on IUG may differ in moderate and heavy smokers because the metabolism of tobacco smoke is affected by factors such as duration, dose, and timing of exposure (Melikian, Djordjevic, Chen, Richie, & Stellman, 2007). Higher doses or more frequent exposure may alter the balance between the participation of *NQO1* in activation versus detoxification reactions, which likely determines its net contribution to cellular toxicity. The apparent protective effect that was observed for the low-activity Ser allele suggests that at moderate cigarette exposure, the effects of fetal *NQO1* on IUG predominantly reflect its catalysis of Phase I reactions rather than Phase II reactions. The loss of the protective effect in the offspring of heavy smokers might be biochemically explained by a dose-related change in the catalytic activity of *NQO1* toward greater participation in detoxification reactions (Xu et al., 2001). This pattern of results, which has been observed in several studies of cancer epidemiology, conforms to a theory of disease susceptibility that proposes that genetic factors are more important in individuals with moderate smoke exposure than in heavily exposed individuals for whom the environmental risk is overwhelming (Vineis, 2002). However, the authors are conscious of the challenges in interpreting their epidemiological findings in mechanistic terms. Additional exploration in vitro will be necessary to elucidate the functional relation between dose and frequency of exposure and xenobiotic metabolism.

Both the direction of this result and the implied mechanism are consistent with evidence from a Japanese sample that found a protective effect of the maternal Ser allele against the effects of smoking to retard fetal growth (Sasaki et al., 2008). In this study of Japanese families, children born to mothers possessing one or more *NQO1* Ser alleles and who smoked during pregnancy did not suffer significantly lower birth weight than children born to nonsmoking mothers of the same genotype. In contrast, children whose mothers were homozygous for the wild type Pro allele and smoked during pregnancy were born on average 199 g lighter than children whose mothers were homozygous for the Pro allele and who did not smoke during pregnancy (interaction $p < .05$). In addition to its likely interactive effects with tobacco smoke exposure on fetal growth, the Ser allele has been found to protect against respiratory distress in response to airborne pollutants (Bergamaschi et al., 2001; Corradi et al., 2002; David et al., 2003; Millstein, Conti, Gilliland, & Gauderman, 2006).

These studies provide convergent evidence for the importance for developmental and health outcomes of the role of *NQO1* in metabolizing toxic pollutants. However, the magnitude and direction of the association may differ depending on the specific exposure, disease, and population under study. For example, homozygotes for the variant allele are at *increased* risk of bone marrow toxicity after benzene exposure (Nebert, Roe, Vandale, Bingham, & Oakley, 2002). Furthermore, the effects of the *NQO1* Pro187Ser allele may differ depending on ethnic background. The variant allele, which is considerably more common in Asian than in Caucasian populations (Ross & Siegel, 2004), is associated with increased risk of cancers of the lung, bladder, and lower intestine in White populations but may be associated with reduced risk of lung cancer in Asian populations (Chao, Zhang, Berthiller, Boffetta, & Hashibe, 2006; Kiyohara, Yoshimasu, Takayama, & Nakanishi, 2005). The elevation in risk of lung cancer associated with the *NQO1* Ser allele in Caucasians may be further moderated by tobacco use (Chao et al., 2006).

Xenobiotic metabolism by fetal cells may moderate the strength of association between intrauterine smoke exposure and other childhood cognitive and behavioral outcomes, especially those that are robustly associated with low birth weight (for reviews, see Breslau, 1995; Bhutta, Cleves, Casey, Cradock, & Anand, 2002). However, the evidence for a causal relation between maternal smoking during pregnancy and children's cognitive and behavioral outcomes is equivocal (D'Onofrio et al., 2008).

Genetic differences within twin pairs were associated with differences in fetal growth among twins who were exposed to tobacco in utero. Testing for the effects of fetal genotype within families avoids many potential confounds, including gestational and familial effects. In particular, the research design obviates the confound between fetal genotype and maternal genotype, circumvents the effects of maternal smoking on gestational age, and is robust to population stratification. However, the circumstances of gestation and birth are substantially different for twins than for singletons, who account for around 99% of live births. Twins have shorter gestations and lower birth weight than singletons. Using a measure of fetal growth corrected for gestational age may increase the likelihood that the results will generalize to singletons (Morley, 2005). Although the shorter gestation and crowding in the uterus restrict fetal growth in twins relative to singletons, this restriction of range should make it harder rather than easier to detect differences in fetal growth due to maternal smoking or fetal genotype.

Although well suited to testing hypotheses about the effects of fetal genotype within families, the current study loses scope by not genotyping the parents. The maternal genotype directly influences perinatal outcomes over and above its contribution to the fetal genome, as has been demonstrated by the greater risk ratio for being born SGA associated with a family history of SGA in the maternal line than in the paternal line (Svensson, Pawitan, Cnattingius, Reilly, & Lichtenstein, 2006). The current study has tested marginal associations between fetal genotype and IUG, but there may also be interactive effects between maternal genotype and fetal genotype on IUG. In addition, genotyping both parents in addition to the children would allow us to test hypotheses about parent-of-origin effects relating to the fetal genotype (Infante-Rivard & Weinberg, 2005).

The measures of infant birth weight, gestational age, and maternal smoking were obtained by retrospective self-report. More than a dozen studies have assessed the validity of maternally reported birth weight (reviewed in Adegboye & Heitmann, 2008; Lucia, Luo, Gardiner, Paneth, & Breslau, 2006). In these studies from diverse populations, the correlations between maternal reports and documentary records of birth weight range from 0.85 to 0.98 (Lucia et al., 2006). These studies therefore suggest that between 0.85 and 0.98 of the variability in parental reports is shared with documented birth weight. The largest sample compared parentally reported birth weight to birth certificates for over 46,000 children and found that 89% of reported birth weights were accurate to within 29 g (Gayle, Yip, Frank, Nieburg, & Binkin, 1988). The proportion of accurately recalled birth weights was the same for children aged 1 year or older as for younger children. A more recent sample of more than 11,000 British children aged around 9 months reported recall accuracy of 82% within 30 g and 92% within 100 g of the registered birth weight (Tate, Dezateux, Cole, & Davidson, 2005). Given that birth weight differences between smoking and nonsmoking groups in the sample and within-twin pairs were well over 100 g, the authors believe that the effects of inaccurate parental recall of birth weight are likely to be small in comparison with the effects of smoking and genotype that are reported here. Gestational age is also generally accurately recalled by mothers, although the research base is smaller (Adegboye & Heitmann, 2008). In the current study, the results of the analyses within families are in any case unaffected by inaccurate measurement of gestational age.

Maternal smoking was also measured by retrospective self-report. It is likely that pregnant women underreport tobacco use, although accurate measures of tobacco intake can be obtained by these means (Kesmodel & Olsen, 1999). In the current study, dose-related decreases in birth weight provide evidence of the validity of both the measure of tobacco intake and the measure of birth weight. The possibility that inaccurate recall may have contributed to variability in both variables, leading to underestimation of the true effect sizes, cannot be ruled out. There is, in addition, some doubt as to whether retrospective self-reports can accurately assess the timing of smoking cessation among mothers who claimed not to have smoked throughout their pregnancy (Heath et al., 2003). Although the authors are conscious of the importance of questions concerning when mothers smoked during pregnancy in relation to perinatal risk, they believe that the current data set does not provide an adequate basis for addressing these issues.

No significant associations were found between fetal polymorphisms in the folate pathways and IUG. The investigation of the relation between folate genotype and fetal growth, however, lacked information on maternal diet during pregnancy. Although this would not confound the genetic association with IUG within families, it is possible that the effects of genetic association are strongest in families in which maternal intake of folic acid is low (Nurk et al., 2004). If true, this possibility would reduce the power of this investigation to detect the effects of folate pathway genes on IUG. It is likely that mothers with low folate intake were represented in this sample because of the absence of mandatory folic acid supplementation in the United Kingdom. The authors therefore believe that these negative results do provide some evidence against the influence of fetal polymorphisms in the folate pathway on IUG, consistent with previous investigations of fetal *MTHFR* and *SHMT1* genotype (Glanville et al., 2006; Infante-Rivard et al., 2002; Relton, Pearce, Burn, et al., 2005) but *contra* a marginally significant report of association between fetal A66G *MTRR* (rs1801394) genotype and growth restriction in babies born at term (Glanville et al., 2003).

Finally, the genetic variants that were studied may moderate the risk of environmental exposures other than smoking that potentially also affect birth weight, such as diesel exhaust exposure and folate intake. A priority in this sample, therefore, will be to conduct further genotyping in twins whose mothers did not smoke during pregnancy, in order to allow conventional testing of a hypothesis of gene–environment interaction. Future epidemiological research should investigate the replicability of this result in other populations and in singletons, the influence of maternal genotype, the possibility of Gene × Gene interactions (i.e., epistasis), and imprinting.

References

Abecasis, G. R., Cardon, L. R., & Cookson, W. O. C. (2000). A general test of association for quantitative traits in nuclear families. *American Journal of Human Genetics*, *66*, 279–292.

Adegboye, A. R. A., & Heitmann, B. (2008). Accuracy and correlates of maternal recall of birthweight and gestational age. *BJOG: An International Journal of Obstetrics and Gynaecology*, *115*, 886–893.

Aylward, G. P. (2005). Neurodevelopmental outcomes of infants born prematurely. *Journal of Developmental and Behavioral Pediatrics*, *26*, 427–440.

Bartsch, H., Nair, U., Risch, A., Rojas, M., Wikman, H., & Alexandrov, K. (2000). Genetic polymorphism of CYP genes, alone or in combination, as a risk modifier of tobacco-related cancers. *Cancer Epidemiology Biomarkers & Prevention*, *9*, 3–28.

Bergamaschi, E., De Palma, G., Mozzoni, P., Vanni, S., Vettori, M. V., Broeckaert, F., et al. (2001). Polymorphism of quinone-metabolizing enzymes and susceptibility to ozone-induced acute effects. *American Journal of Respiratory and Critical Care Medicine*, *163*, 1426–1431.

Bhutta, A. T., Cleves, M. A., Casey, P. H., Cradock, M. M., & Anand, K. J. S. (2002). Cognitive and behavioral outcomes of school-aged children who were born preterm—A meta-analysis. *JAMA: The Journal of the American Medical Association*, *288*, 728–737.

Bourgain, C., Abney, M., Schneider, D., Ober, C., & McPeek, M. S. (2004). Testing for Hardy–Weinberg equilibrium in samples with related individuals. *Genetics*, *168*, 2349–2361.

Breslau, N. (1995). Psychiatric sequelae of low birth weight. *Epidemiologic Reviews*, *17*, 96–106.

Cadenas, E. (1995). Antioxidant and prooxidant functions of DT-diaphorase in quinone metabolism. *Biochemical Pharmacology*, *49*, 127–140.

Cartlidge, P. H. T., & Stewart, J. H. (1995). Effect of changing the stillbirth definition on evaluation of perinatal mortality rates. *Lancet*, *346*, 486–488.

Chao, C., Zhang, Z. F., Berthiller, J., Boffetta, P., & Hashibe, M. (2006). NAD(P)H: Quinone oxidoreductase 1 (NQO1) Pro187Ser polymorphism and the risk of lung, bladder, and colorectal cancers: A meta-analysis. *Cancer Epidemiology Biomarkers & Prevention*, *15*, 979–987.

Charles, D. H. M., Ness, A. R., Campbell, D., Smith, G. D., Whitley, E., & Hall, M. H. (2005). Folic acid supplements in pregnancy and birth outcome: Re-analysis of a large randomised controlled trial and update of Cochrane review. *Paediatric and Perinatal Epidemiology*, *19*, 112–124.

Chen, D. F., Hu, Y. H., Yang, F., Li, Z. P., Wu, B. Y., Fang, Z., et al. (2005). Cytochrome P450 gene polymorphisms and risk of low birth weight. *Genetic Epidemiology, 28,* 368–375.

Corradi, M., Alinovi, R., Goldoni, M., Vettori, M. V., Folesani, G., Mozzoni, P., et al. (2002). Biomarkers of oxidative stress after controlled human exposure to ozone. *Toxicology Letters, 134,* 219–225.

David, G. L., Romieu, I., Sienra-Monge, J. J., Collins, W. J., Ramirez-Aguilar, M., del Rio-Navarro, B. E., et al. (2003). Nicotinamide adenine dinucleotide (phosphate) reduced: Quinone oxidoreductase and glutathione S-transferase M1 polymorphisms and childhood asthma. *American Journal of Respiratory and Critical Care Medicine, 168,* 1199–1204.

De Flora, S., Bennicelli, C., Dagostini, F., Izzotti, A., & Camoirano, A. (1994). Cytosolic activation of aromatic and heterocyclic amines—Inhibition by dicoumarol and enhancement in viral hepatitis B. *Environmental Health Perspectives, 102,* 69–74.

D'Onofrio, B. M., Hulle, C. A., Waldman, I. D., Rodgers, J. L., Harden, K. P., Rathouz, P. J., et al. (2008). Smoking during pregnancy and offspring externalizing problems: An exploration of genetic and environmental confounds. *Development and Psychopathology, 20,* 139–164.

Engel, S. M., Olshan, A. F., Siega-Riz, A. M., Savitz, D. A., & Chanock, S. J. (2006). Polymorphisms in folate metabolizing genes and risk for spontaneous preterm and small-for-gestational age birth. *American Journal of Obstetrics and Gynecology, 195,* 1231–1239.

Franchi, F., Cetin, I., Todros, T., Antonazzo, P., De Santis, M. S. N., Cardaropoli, S., et al. (2004). Intrauterine growth restriction and genetic predisposition to thrombophilia. *Haematologica, 89,* 444–449.

Fredriksen, A., Meyer, K., Ueland, P. M., Vollset, S. E., Grotmol, T., & Schneede, J. (2007). Large-scale population-based metabolic phenotyping of thirteen genetic polymorphisms related to one-carbon metabolism. *Human Mutation, 28,* 856–865.

Freeman, B., Powell, J., Ball, D., Hill, L., Craig, I., & Plomin, R. (1997). DNA by mail: An inexpensive and noninvasive method for collecting DNA samples from widely dispersed populations. *Behavior Genetics, 27,* 251–257.

Fulker, D. W., Cherny, S. S., Sham, P. C., & Hewitt, J. K. (1999). Combined linkage and association sib-pair analysis for quantitative traits. *American Journal of Human Genetics, 64,* 259–267.

Gayle, H. D., Yip, R., Frank, M. J., Nieburg, P., & Binkin, N. J. (1988). Validation of maternally reported birth weights among 46,637 Tennessee WIC program participants. *Public Health Reports, 103,* 143–147.

Gebhardt, G. S., Scholtz, C. L., Hillermann, R., & Odendaal, H. J. (2001). Combined heterozygosity for methylenetetrahydrofolate reductase (MTHFR) mutations C677T and A1298C is associated with abruptio placentae but not with intrauterine growth restriction. *European Journal of Obstetrics Gynecology and Reproductive Biology, 97,* 174–177.

Glanville, T., Yates, Z., Hopkins, S., Walker, J. J., Lucock, M., & Simpson, N. (2003). Growth restriction is associated with fetal carriage of A66G methionine synthase reductase SNP. *American Journal of Obstetrics and Gynecology, 189,* S226.

Glanville, T., Yates, Z., Ovadia, L., Walker, J. J., Lucock, M., & Simpson, N. A. B. (2006). Fetal folate C677T methylenetetrahydrofolate reductase gene polymorphism and low birth weight. *Journal of Obstetrics and Gynaecology, 26,* 11–14.

Goldenberg, R. L., Tamura, T., Cliver, S. P., Cutter, G. R., Hoffman, H. J., & Copper, R. L. (1992). Serum folate and fetal growth-retardation—A matter of compliance. *Obstetrics and Gynecology, 79,* 719–722.

Hajos, A. K. D., & Winston, G. W. (1991). Purified NAD(P)H-quinone oxidoreductase enhances the mutagenicity of dinitropyrenes in vitro. *Journal of Biochemical Toxicology, 6,* 277–282.

Health Development Agency. (2004). *The evidence of effectiveness of public health interventions—and the implications.* London: Author.

Heath, A. C., Knopik, V. S., Madden, P. A., Neuman, R. J., Lynskey, M. J., Slutske, W. S., et al. (2003). Accuracy of mothers' retrospective reports of smoking during pregnancy: Comparison with twin sister informant ratings. *Twin Research, 6,* 297–301.

Infante-Rivard, C., Rivard, G. E., Yotov, W. V., Genin, E., Guiguet, M., Weinberg, C., et al. (2002). Absence of association of thrombophilia polymorphisms with intrauterine growth restriction. *New England Journal of Medicine, 347,* 19–25.

Infante-Rivard, C., & Weinberg, C. R. (2005). Parent-of-origin transmission of thrombophilic alleles to intrauterine growth restricted newborns and transmission-ratio distortion in unaffected newborns. *American Journal of Epidemiology, 162,* 891–897.

Infante-Rivard, C., Weinberg, C. R., & Guiguet, M. (2006). Xenobiotic-metabolizing genes and small-for-gestational-age births—Interaction with maternal smoking. *Epidemiology, 17,* 38–46.

Jensen, L. E., Hoess, K., Whitehead, A. S., & Mitchell, L. E. (2005). The NAT1 C1095A polymorphism, maternal multivitamin use and smoking, and the risk of spina bifida. *Birth Defects Research: Part A. Clinical and Molecular Teratology, 73,* 512–516.

Kato, I., Dnistrian, A. M., Schwartz, M., Toniolo, P., Koenig, K., Shore, R. E., et al. (1999). Epidemiologic correlates of serum folate and homocysteine levels among users and non-users of vitamin supplement. *International Journal for Vitamin and Nutrition Research, 69,* 322–329.

Kesmodel, U., & Olsen, S. F. (1999). Smoking habits among pregnant Danish women: Reliability of information recorded after delivery. *Journal of Epidemiology and Community Health, 53,* 239–242.

Kim, Y. J., Hong, Y. C., Lee, K. H., Park, H. J., Park, E. A., Moon, H. S., et al. (2005). Oxidative stress in pregnant women and birth weight reduction. *Reproductive Toxicology, 19,* 487–492.

Kiyohara, C., Yoshimasu, K., Takayama, K., & Nakanishi, Y. (2005). NQO1, MPO, and the risk of lung cancer: A HuGE review. *Genetics in Medicine, 7*, 463–478.

Knox, R. J., Boland, M. P., Friedlos, F., Coles, B., Southan, C., & Roberts, J. J. (1988). The nitroreductase enzyme in Walker cells that activates 5-(aziridin-1-yl)-2,4-dinitrobenzamide (CB 1954) to 5-(aziridin-1-yl)-4-hydroxylamino-2-nitrobenzamide is a form of NAD(P)H dehydrogenase (quinone) (EC 1.6.99.2). *Biochemical Pharmacology, 37*, 4671–4677.

Kraft, P., Yen, Y. C., Stram, D. O., Morrison, J., & Gauderman, W. J. (2007). Exploiting gene–environment interaction to detect genetic associations. *Human Heredity, 63*, 111–119.

Kramer, M. S. (2003). The epidemiology of adverse pregnancy outcomes: An overview. *Journal of Nutrition, 133*, 1592S–1596S.

Kupferminc, M. J., Eldor, A., Steinman, N., Many, A., Bar-Am, A., Jaffa, A., et al. (1999). Increased frequency of genetic thrombophilia in women with complications of pregnancy. *New England Journal of Medicine, 340*, 9–13.

Li, Y., & Wang, H. (2004). In utero exposure to tobacco and alcohol modifies neurobehavioral development in mice offspring: Consideration a role of oxidative stress. *Pharmacological Research, 49*, 467–473.

Lucia, V. C., Luo, Z., Gardiner, J. C., Paneth, N., & Breslau, N. (2006). Reports of birthweight by adolescents and their mothers: Comparing accuracy and identifying correlates. *Paediatric and Perinatal Epidemiology, 20*, 520–527.

Lumley, J., Oliver, S., Chamberlain, C., & Oakley, L. (1998). Interventions for promoting smoking cessation during pregnancy. *Cochrane Database of Systematic Reviews, 3*, CD001055.

Lumley, J., Watson, L., Watson, M., & Bower, C. (1998). Periconceptional supplementation with folate and/or multivitamins for preventing neural tube defects. *Cochrane Database of Systematic Reviews, 4*, CD001056.

Martinelli, P., Grandone, E., Colaizzo, D., Paladini, D., Scianname, N., Margaglione, M., et al. (2001). Familial thrombophilia and the occurrence of fetal growth restriction. *Haematologica, 86*, 428–431.

McCowan, L. M. E., Craigie, S., Taylor, R. S., Ward, C., McLintock, C., & North, R. A. (2003). Inherited thrombophilias are not increased in "idiopathic" small-for-gestational-age pregnancies. *American Journal of Obstetrics and Gynecology, 188*, 981–985.

McDonald, S. D., Perkins, S. L., Jodouin, C. A., & Walker, M. C. (2002). Folate levels in pregnant women who smoke: An important gene/environment interaction. *American Journal of Obstetrics and Gynecology, 187*, 620–625.

Melikian, A. A., Djordjevic, M. V., Chen, S., Richie, J., & Stellman, S. D. (2007). Effect of delivered dosage of cigarette smoke toxins on the levels of urinary biomarkers of exposure. *Cancer Epidemiology Biomarkers & Prevention, 16*, 1408–1415.

Millstein, J., Conti, D. V., Gilliland, F. D., & Gauderman, W. J. (2006). A testing framework for identifying susceptibility genes in the presence of epistasis. *American Journal of Human Genetics, 78*, 15–27.

Mitchell, E. A., Robinson, E., Clark, P. M., Becroft, D. M. O., Glavish, N., Pattison, N. S., et al. (2004). Maternal nutritional risk factors for small for gestational age babies in a developed country: A case-control study. *Archives of Disease in Childhood, 89*, F431–F435.

Moffitt, T. E., Caspi, A., & Rutter, M. (2005). Strategy for investigating interactions between measured genes and measured environments. *Archives of General Psychiatry, 62*, 473–481.

Morley, R. (2005). Can we generalise from findings in twins? *Paediatric and Perinatal Epidemiology, 19*, 54–59.

Murphy, R. P., Donoghue, C., Nallen, R. J., D'Mello, M., Regan, C., Whitehead, A. S., et al. (2000). Prospective evaluation of the risk conferred by factor V Leiden and thermolabile methylenetetrahydrofolate reductase polymorphisms in pregnancy. *Arteriosclerosis Thrombosis and Vascular Biology, 20*, 266–270.

Myatt, L., & Cui, X. L. (2004). Oxidative stress in the placenta. *Histochemistry and Cell Biology, 122*, 369–382.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2003). *Mx: Statistical modeling* (6th ed.). Richmond, VA: Department of Psychiatry.

Nebert, D. W., Roe, A. L., Vandale, S. E., Bingham, E., & Oakley, G. G. (2002). NAD(P)H : Quinone oxidoreductase (NQ01) polymorphism, exposure to benzene, and predisposition to disease: A HuGE review. *Genetics in Medicine, 4*, 62–70.

Neggers, Y. H., Goldenberg, R. L., Tamura, T., Cliver, S. P., & Hoffman, H. J. (1997). The relationship between maternal dietary intake and infant birthweight. *Acta Obstetricia Et Gynecologica Scandinavica, 76*, 71–75.

Nurk, E., Tell, G. S., Refsum, H., Ueland, P. M., & Vollset, S. E. (2004). Associations between maternal methylenetetrahydrofolate reductase polymorphisms and adverse outcomes of pregnancy: The Hordaland Homocysteine Study. *American Journal of Medicine, 117*, 26–31.

Park, H., Kim, Y. J., Ha, E. H., Lee, B. E., Park, B. H., Lee, H. Y., et al. (2005). Interaction between maternal serum folate and the methylenetetrahydrofolate reductase (MTHFR) polymorphisms on infant birthweight. *Molecular & Cellular Toxicology, 1*, 130–136.

Pollack, H., Lantz, P. M., & Frohna, J. G. (2000). Maternal smoking and adverse birth outcomes among singletons and twins. *American Journal of Public Health, 90*, 395–400.

Posthuma, D., de Geus, E. J. C., Boomsma, D. I., & Neale, M. C. (2004). Combined linkage and association tests in Mx. *Behavior Genetics, 34*, 179–196.

Price, T. S. (2003). *Genetic and environmental influences on verbal and nonverbal domains of cognitive ability and low cognitive ability: A study of 2-, 3-, and 4-year-old twins.* London: Institute of Psychiatry, King's College, University of London.

Price, T. S., Freeman, B., Craig, I., Petrill, S. A., Ebersole, L., & Plomin, R. (2000). Infant zygosity can be assigned by parental report questionnaire data. *Twin Research, 3*, 129–133.

R Development Core Team. (2007). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Rao, S., Yajnik, C. S., Kanade, A., Fall, C. H. D., Margetts, B. M., Jackson, A. A., et al. (2001). Intake of micronutrient-rich foods in rural Indian mothers is associated with the size of their babies at birth: Pune maternal nutrition study. *Journal of Nutrition, 131*, 1217–1224.

Refsum, H., Nurk, E., Smith, A. D., Ueland, P. M., Gjesdal, C. G., Bjelland, I., et al. (2006). The Hordaland Homocysteine Study: A community-based study of homocysteine, its determinants, and associations with disease. *Journal of Nutrition, 136*, 1731S–1740S.

Relton, C. L., Pearce, M. S., Burn, J., & Parker, L. (2005). An investigation of folate-related genetic factors in the determination of birthweight. *Paediatric and Perinatal Epidemiology, 19*, 360–367.

Relton, C. L., Pearce, M. S., & Parker, L. (2005). The influence of erythrocyte folate and serum vitamin B-12 status on birth weight. *British Journal of Nutrition, 93*, 593–599.

Rennert, H., Zeigler-Johnson, C. M., Addya, K., Finley, M. J., Walker, A. H., Spangler, E., et al. (2005). Association of susceptibility alleles in ELAC2/HPC2, RNASEL/HPC1, and MSR1 with prostate cancer severity in European American and African American men. *Cancer Epidemiology Biomarkers & Prevention, 14*, 949–957.

Rooney, R., Hay, D., & Levy, F. (2003). Small for gestational age as a predictor of behavioral and learning problems in twins. *Twin Research, 6*, 46–54.

Ross, D., & Siegel, D. (2004). NAD(P)H: Quinone oxidoreductase 1 (NQO1, DT-diaphorase), functions and pharmacogenetics. In H. Sies & L. Packer (Eds.), *Quinones and quinone enzymes, Part B* (Vol. 382, pp. 115–144). Amsterdam: Elsevier Academic Press.

Salihu, H. M., Aliyu, M. H., & Kirby, R. S. (2005). In utero nicotine exposure and fetal growth inhibition among twins. *American Journal of Perinatology, 22*, 421–427.

Salomon, O., Seligsohn, U., Steinberg, D. M., Zalel, Y., Lerner, A., Rosenberg, N., et al. (2004). The common prothrombotic factors in nulliparous women do not compromise blood flow in the feto-maternal circulation and are not associated with preeclampsia or intrauterine growth restriction. *American Journal of Obstetrics and Gynecology, 191*, 2002–2009.

Sasaki, S., Kondo, T., Sata, F., Saijo, Y., Katoh, S., Nakajima, S., et al. (2006). Maternal smoking during pregnancy and genetic polymorphisms in the Ah receptor, CYP1A1 and GSTM1 affect infant birth size in Japanese subjects. *Molecular Human Reproduction, 12*, 77–83.

Sasaki, S., Sata, F., Katoh, S., Saijo, Y., Nakajima, S., Washino, N., et al. (2008). Adverse birth outcomes associated with maternal smoking and polymorphisms in the N-nitrosamine-metabolizing enzyme genes NQO1 and CYP2E1. *American Journal of Epidemiology, 167*, 6.

Siegel, D., Anwar, A., Winski, S. L., Kepa, J. K., Zolman, K. L., & Ross, D. (2001). Rapid polyubiquitination and proteasomal degradation of a mutant form of NAD(P)H: Quinone oxidoreductase 1. *Molecular Pharmacology, 59*, 263–268.

Siegel, D., McGuinness, S. M., Winski, S. L., & Ross, D. (1999). Genotype–phenotype relationships in studies of a polymorphism in NAD(P)H: Quinone oxidoreductase 1. *Pharmacogenetics, 9*, 113–121.

Sram, R. J., Binkova, B., Dejmek, J., Chvatalova, I., Solansky, N., & Topinka, J. (2006). Association of DNA adducts and genotypes with birth weight. *Mutation Research-Genetic Toxicology and Environmental Mutagenesis, 608*, 121–128.

Sram, R. J., Binkova, B., Lnenickova, Z., Solansky, I., & Dejmek, J. (2005). The impact of plasma folate levels of mothers and newborns on intrauterine growth retardation and birth weight. *Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis, 591*, 302–310.

Stark, K. D., Pawlosky, R. J., Sokol, R. J., Hannigan, J. H., & Salem, N. (2007). Maternal smoking is associated with decreased 5-methyltetrahydrofolate in cord plasmas. *American Journal of Clinical Nutrition, 85*, 796–802.

Steegers-Theunissen, R. P., Van Iersel, C. A., Peer, P. G., Nelen, W. L., & Steegers, E. A. (2004). Hyperhomocysteinemia, pregnancy complications, and the timing of investigation. *Obstetrics and Gynecology, 104*, 336–343.

Svensson, A. C., Pawitan, Y., Cnattingius, S., Reilly, M., & Lichtenstein, P. (2006). Familial aggregation of small-for-gestational-age births: The importance of fetal genetic effects. *American Journal of Obstetrics and Gynecology, 194*, 475–479.

Tate, A. R., Dezateux, C., Cole, T. J., & Davidson, L. (2005). Factors affecting a mother's recall of her baby's birth weight. *International Journal of Epidemiology, 34*, 688–695.

Traver, R. D., Siegel, D., Beall, H. D., Phillips, R. M., Gibson, N. W., Franklin, W. A., et al. (1997). Characterization of a polymorphism in NAD(P)H: Quinone oxidoreductase (DT-diaphorase). *British Journal of Cancer, 75*, 69–75.

Trouton, A., Spinath, F. M., & Plomin, R. (2002). Twins Early Development Study (TEDS): A multivariate, longitudinal genetic investigation of language, cognition and behavior problems in childhood. *Twin Research, 5*, 444–448.

Verspyck, E., Le Cam-Duchez, V., Goffinet, F., Tron, F., Marpeau, L., & Borg, J. Y. (2002). Thrombophilia and immunological disorders in pregnancies as risk factors for small for gestational age infants. *BJOG: An International Journal of Obstetrics and Gynaecology, 109*, 28–33.

Vineis, P. (2002). The relationship between polymorphisms of xenobiotic metabolizing enzymes and susceptibility to cancer. *Toxicology, 181, 182*, 457–462.

Walmsley, C. M., Bates, C. J., Prentice, A., & Cole, T. J. (1999). Relationship between cigarette smoking and nutrient intakes and blood status indices of older people living in the UK: Further analysis of data from the National Diet and Nutrition Survey of people aged 65 years and over, 1994/95. *Public Health Nutrition, 2,* 199–208.

Wang, X. B., Zuckerman, B., Pearson, C., Kaufman, G., Chen, C. Z., Wang, G. Y., et al. (2002). Maternal cigarette smoking, metabolic gene polymorphism, and infant birth weight. *JAMA: The Journal of the American Medical Association, 287,* 195–202.

Wu, T., Hu, Y. H., Chen, C. Z., Yang, F., Li, Z. P., Fang, Z., et al. (2007). Passive smoking, metabolic gene polymorphisms, and infant birth weight in a prospective cohort study of Chinese women. *American Journal of Epidemiology, 166,* 313–322.

Xu, L. L., Wain, J. C., Miller, D. P., Thurston, S. W., Su, L., Lynch, T. J., et al. (2001). The NAD(P)H:Quinone oxidoreductase 1 gene polymorphism and lung cancer: Differential susceptibility based on smoking behavior. *Cancer Epidemiology Biomarkers & Prevention, 10,* 303–309.

Child Development, January/February 2010, Volume 81, Number 1, Pages 115–130

Prenatal Antecedents of Newborn Neurological Maturation

Janet A. DiPietro, Katie T. Kivlighan, Kathleen A. Costigan, Suzanne E. Rubin,
Dorothy E. Shiffler, Janice L. Henderson, and Joseph P. Pillion
Johns Hopkins University

Fetal neurobehavioral development was modeled longitudinally using data collected at weekly intervals from 24 to 38 weeks gestation in a sample of 112 healthy pregnancies. Predictive associations between 3 measures of fetal neurobehavioral functioning and their developmental trajectories to neurological maturation in the first weeks after birth were examined. Prenatal measures included fetal heart rate (FHR) variability, fetal movement, and coupling between fetal motor activity and heart rate patterning; neonatal outcomes include a standard neurologic examination (n = 97) and brainstem auditory evoked potential (BAEP; n = 47). Optimality in newborn motor activity and reflexes was predicted by fetal motor activity, FHR variability, and somatic–cardiac coupling predicted BAEP parameters. Maternal pregnancy-specific psychological stress was associated with accelerated neurologic maturation.

> If we pursue our quest beyond the newborn period, we find ourselves suddenly in an entirely new situation, where our organism is not seen, nor scarcely felt nor heard. Our environmental situation has become, if not at once more complex, at least to be likened to the postnatal environment only with great difficulty. (Sontag & Richards, 1938, p. 1)

While the last decade has been marked by increasing appreciation of the role of the prenatal environment in providing the substrate for postnatal health and development, recognition of the continuous nature of human development from conception was evident in the earliest publications of the Fels Research Institute, established in 1929. Since that time, many of the core interests of developmental psychology concerning development and expression of individual differences and the moderating role of early environmental influences have begun to converge with epidemiologic methodology. The construct of ''fetal programming'' has been applied broadly to represent discoveries of prenatal influences on postnatal functioning, typically in adulthood (Barker, 2006; O'Brien, Wheeler, & Barker, 1999; Young, 2002). Although this approach has generated an enormous body of data, thereby sparking great interest in the prenatal period, most studies rely on readily available data sources, such as birth weight, that provide only vague proxy for the gestational environment and can offer little information about mechanisms mediating observed associations.

A second, more direct, approach is to measure function during the prenatal period to evaluate its role as the foundation for postnatal function. This permits measurement of a single construct and allows determination of how early experiences or exposures might affect the latter via their influence on the former. Technological advances available only after the dissolution of the Fels Institute have made clear that by the end of gestation developmental parameters that are measured extensively in the neonate and infant, and are integral to theories of development, originate at neither term nor with birth (Als, 1982; Prechtl, 1984). The view that fetal neurobehaviors reflect neurological development has been supported by studies conducted in healthy populations (Amiel-Tison, Gosselin, & Kurjak, 2006; DiPietro, Irizarry, Hawkins, Costigan,

---

This research was supported by NIH/NICHD Grant R01 HD27592 to the first author. Newborn evoked potential data collection was supported by a research grant equipment loan from Intelligent Hearing Systems, Miami, FL. We thank Rafael Irizarry, Department of Biostatistics, for his contribution to the original design of the fetal protocol, and Kristen Byrnes, for her assistance with Dubowitz training and administration. We remain indebted to the diligent and generous support of our study families, without which this work would not be possible.

Correspondence concerning this article should be addressed to Janet A. DiPietro, Department of Population, Family and Reproductive Health, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe St., W1033 Baltimore, MD 21205. Electronic mail may be sent to jdipietr@jhsph.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8101-0008

& Pressman, 2001; Hepper, 1995; Krasnegor et al., 1998; Nijhuis & ten Hof, 1999; Sandman, Wadhwa, Hetrick, Porto, & Peeke, 1997). Further support is provided by observations of differences in neurobehavioral functioning in fetuses afflicted by congenital anomalies related to the nervous system (Hepper & Shahidullah, 1992; Horimoto et al., 1993; Maeda et al., 2006; Romanini & Rizzo, 1995); exposure to deleterious antenatal conditions that affect development, including growth restriction (Nijhuis et al., 2000) and maternal diabetes (Kainer, Prechtl, Engele, & Einspieler, 1997); and exposure to potentially neurotoxic substances (Gingras & O'Donnell, 1998; Mulder, Morssink, van der Schee, & Visser, 1998). The first aim of this study was to evaluate the association between central nervous system maturation in the fetus, as indicated by fetal neurobehavioral development, and neurologic functioning of the newborn infant in an effort to identify early indicators of optimal function.

While we have examined a number of fetal neurobehavioral parameters in past cohorts, three were selected for this analysis based on conceptual considerations regarding relevance to postnatal neurological functioning. The first parameter is spontaneous variability in fetal heart rate (FHR). Measurement of phasic or nonphasic cardiac variability has had a distinguished history in developmental science as a marker of the physiological regulation that corresponds to infant and child performance and behavior. Variability in FHR is among the most prominently accessible features of the fetus, and is central to clinical antepartum assessment as an indicator of the developing balance between parasympathetic and sympathetic innervation (Freeman, Garite, & Nageotte, 1991). Recently, FHR variability at or after 28 weeks gestation and steeper developmental trajectories were shown to be significantly associated with mental, psychomotor, and language development in the 3rd year of life (DiPietro, Bornstein, Hahn, Costigan, & Achy-Brou, 2007). Significant stability within individual fetuses during gestation was also demonstrated (DiPietro et al., 2007).

Motor activity is the second prenatal parameter, selected for its conspicuous nature as an individual difference in both the fetus and child (Eaton, McKeen, & Campbell, 2001). A few studies on small samples have suggested conservation of very specific attributes of fetal motor patterning between the fetus and infant (Almli, Ball, & Wheeler, 2001; Groome et al., 1999). Fetal motor activity is significantly associated with activity levels in early childhood for boys, and predictive of a range of regulatory temperament characteristics (DiPietro et al., 2002). The putative assumption is that fetal motor activity prepares the neurologic circuitry and musculature for postnatal function (Prechtl, 1984). Data generated from animal models further reveal that variation in fetal motor behavior reflects ontogenic adaptation to the intrauterine environment, which in turn fosters subsequent maturation (Smotherman & Robinson, 1987).

The third fetal parameter combines FHR and motor activity into a single measure reflecting the degree of covariation between the two. Observation of synchrony between acceleration in heart rate and motor activity in fetuses is long-standing (Sontag & Richards, 1938). Since then, cardiac–somatic coupling has become implicated as a function of parasympathetic control, which becomes the increasingly prominent influence as gestation advances. The association between movement and heart rate in the fetus has been most often attributed to centrally mediated coactivation of cardiac and somatomotor processes (Johnson, Besinger, Thomas, Strobino, & Niebyl, 1992; Timor-Tritsch, Dierker, Zador, Hertz, & Rosen, 1978; Vintzileos, Campbell, & Nochinson, 1986). Documentation of normative development of this association in several longitudinal samples reveals a predictable progression during gestation marked by increased levels of correspondence and diminished latency between parameters (DiPietro, Hodgson, Costigan, Hilton, & Johnson, 1996; DiPietro et al., 2001; DiPietro et al., 2004).

One of the most frequently investigated sources of maternal environmental variation on the developing fetus has been maternal psychological distress. Most studies examine effects on pregnancy outcomes (i.e., birth weight and gestational age). Although a number have reported significant associations between maternal stress and/or anxiety during pregnancy and shortened gestation or restricted growth, results are not uniform (Alder, Fink, Bitzer, Hosli, & Holzgreve, 2007; Littleton, Breitkopf, & Berenson, 2007). Fewer studies link maternal distress to observed child or infant outcomes; of those that do, most examine consequences on attentional or behavioral regulation (Davis et al., 2004; Gutteling et al., 2005; Huizink, Robles de Medina, Mulder, Visser, & Buitelaar, 2002; Van den Bergh et al., 2005). Examination of child performance on standardized assessments has yielded conflicting results. For example, two studies report deleterious effects between maternal psychological stress and/or anxiety on Bayley Scales of Infant Development (Buitelaar, Huizink, Mulder, Robles de Medina, & Visser, 2003; Laplante et al.,

2004) whereas another reports that maternal psychological distress accelerates development (DiPietro, Novak, Costigan, Atella, & Reusing, 2006). The handful of studies examining the most proximal associations, those between maternal psychological distress and fetal functional activity, are neither consistent nor conclusive. The second aim of this study was to systematically examine the contribution of maternal nonspecific anxiety levels and pregnancy-specific stress on both fetal and neonatal maturation. Boys have typically been observed to be more vulnerable to prenatal exposures in human studies, and a potential mechanism specific to prenatal stress has recently been uncovered in rodent models (Mueller & Bale, 2008). Thus, potential sex differences in any observed associations will be examined.

Growth and development during gestation both reflect maturative processes. Growth is generally defined as an increase in cell mass or number, whereas development refers to differentiation of function. Although growth is relatively straightforward to measure at birth in terms of size, there are a variety of ways that functional maturation may be expressed and measured. As our focus is on neurologic maturation, we selected two different indicators of neurologic function. The first is a traditional neurological examination of the newborn. Assessments of this type originated in the 1960s and contain a fairly standard repertoire of neonatal characteristics, including tone, posture, primitive reflexes, and behavior. The Dubowitz Neurological Exam (Dubowitz, Mercuri, & Dubowitz, 1998), used in this study, was selected because scoring has been validated around the construct of "optimality" such that the exam distinguishes the optimal point along dimensions in which suboptimal performance occupies each end (e.g., hyporesponsivity through hyperresponsivity). However, despite their widespread use, assessments of this type have a number of limitations, including restriction to features of neonatal functioning that can be observed and/or manipulated and fairly subjective scoring based on "stick figure" drawings of response ranges.

As a result, a second approach was included on a subset of infants using measurement of brainstem auditory evoked potentials (BAEPs) as a more objective and precise indicator of neurological maturation. BAEPs have standard clinical application in the newborn period in evaluation of audition. A growing literature has indicated that they are effective in serving as a marker for more general neural maturation. Variation in BAEP parameters has been linked to prenatal lead exposure (Rothenberg, Poblano, & Schnaas, 2000), prenatal exposure to potentially neurotoxic therapeutic drugs (Poblano et al., 2003), transient depression of Apgar scores (Jiang, Xu, Brosi, Shao, & Wilkinson, 2007), iron deficiency anemia (Roncagliolo, Garrido, Walter, Peirano, & Lozoff, 1998), and breast-feeding and formula fortification (Khedr, Farghaly, El-DinAmry, & Osman, 2004; Unay et al., 2004). These results suggest that BAEP may serve as a marker for functional integrity of the nervous system beyond the specific pathway of conduction.

The BAEP consists of a series of three major peaks or waves (i.e., I, III, and V) that are generated in the eighth cranial nerve and brainstem. The neural generators for Waves I and III in humans are the auditory nerve and cochlear nucleus, respectively. The generator for the most positive peak of Wave V is the termination of the fiber tract of the lateral lemniscus, whereas the following negative trough is produced by slow dendritic potentials in the inferior colliculus (Moller & Jannetta, 1982). BAEP responsivity can be detected in some preterm infants as early as 26 weeks gestation but is not typically measurable until 30–32 weeks postconceptional age (Rotteveel, de Graaf, Colon, Stegeman, & Visco, 1987). The absolute latencies of Waves I, III, and V as well as the interpeak latency intervals progressively decrease over the course of development (Amin, Orlando, Dalzell, Merle, & Guillet, 1999). In full-term infants, the BAEP undergoes rapid changes in the first 4 days of life, with the most rapid changes occurring in the first 24 hr (Yamasaki et al., 1991). The later components (i.e., Waves III and V) undergo more marked changes in latency than the most peripheral component due to the earlier maturation of the peripheral nervous system relative to the later maturing central auditory pathways (Montandon, Cao, Engel, & Grajew, 1979). The prolonged latencies in the BAEP in neonates reflect incomplete neural myelination of the fibers in the auditory pathway, reduced axon diameter, and immaturity in synaptic function (Eggermont & Salamy, 1988; Folsom & Wynne, 1987).

Based on the constructs underlying each dependent and independent measure, we expect the following: (a) fetal motor activity will be most predictive of newborn performance on the standard neurological examination, given the focus on postnatal tone, posture, and motor behavior; (b) fetal measures more closely reflective of neural functioning (i.e., cardiac variability and somatic-cardiac coupling) will be most predictive of event-related potentials (BAEP); and (c) maternal psychological distress will accelerate fetal neurobehavioral development and neonatal neurological functioning.

Method

Participants

Participants were 112 self-referred normotensive, nonsmoking women with normally progressing pregnancies carrying singleton fetuses. Accurate dating of the pregnancy, based on early first trimester pregnancy testing or examination and generally confirmed by early ultrasound was required (*M* gestational age at pregnancy detection = 4.8 weeks; *SD* = 1.2). The sample represents a relatively stable population of well educated (*M* years education = 17.2 years, *SD* = 2.1), mature (*M* age = 31.2, *SD* = 4.6), and married (91%) women. Most were non-Hispanic White (79.5%); the remainder were African American (13.4%), or Hispanic or Asian (7.1%). Fifty-seven (51%) of the fetuses were female.

Procedure

Prenatal. To fully represent the gestational age span from 24 to 38 weeks gestation, participants were stratified into three cohorts with staggered entry into the protocol between 24 and 26 weeks gestation and tested in 3-week intervals. That is, data collection for the first cohort proceeded at 24, 27, 30, 33, and 36 weeks; the second at 25, 28, 31, 34, and 37 weeks; and the third at 26, 29, 32, 35, and 38 weeks. Prenatal visits were scheduled at 13:00 or 15:00. On the day of the visit, women were instructed to eat 1.5 hr prior to the visit but not thereafter. Self-report psychosocial questionnaires were completed upon arrival at the laboratory at each visit. A brief ultrasound scan was administered to determine fetal position followed by 50 min of undisturbed fetal recording.

The psychosocial questionnaires included measures of general anxiety and pregnancy-specific stress, both administered at each visit. The former was evaluated by the Spielberger Trait Anxiety Scales (Y-2; STAI; Spielberger, 1983), one of the most commonly used and extensively validated self-administered measures of anxiety. This questionnaire includes twenty 4-point items; items were reversed as necessary and summed such that higher scores indicate greater trait anxiety. Pregnancy-specific stress was assessed by a shortened form of the Pregnancy Experiences Scale (PES). The original version was previously validated (DiPietro, Ghera, Costigan, & Hawkins, 2004) and contains 41 items specific to pregnancy, each rated on 4-point scales and scored in terms of intensity and frequency. The revised version was developed for the recurring and frequent administration of the current design. The PES–Brief includes the 10 most frequently endorsed hassles and uplifts from the full PES, each rated on the original 4-point scale, and averaged. Higher values reflect greater perceived intensity of either negative or positive stressors during pregnancy. Comparable reliability and stability to the original have been demonstrated (DiPietro, Christensen, & Costigan, 2008).

Fetal data were collected using a Toitu (MT320) fetal actocardiograph (Toitu Co., Ltd, Tokyo, Japan). This monitor detects fetal movement (FM) and FHR with a single wide array transabdominal Doppler transducer and processes this signal through a series of autocorrelation techniques. The actograph detects FM by preserving the remaining signal after bandpassing frequency components of the Doppler signal that are associated with FHR and maternal somatic activity. Reliability studies comparing actograph based versus ultrasound visualized FM have found the performance of this monitor to be highly accurate in detecting both fetal motor activity and quiescence (Besinger & Johnson, 1989; DiPietro, Costigan, & Pressman, 1999; Maeda, Tatsumura, & Utsu, 1999).

Fetal data were digitized and analyzed off-line using software developed in our laboratory (GESTATE; James Long Company, Caroga Lake, NY). Fetal heart rate data underwent error rejection procedures based on moving averages of acceptable values as needed. Variable extraction included FHR variability (standard deviation of each 1-min epoch of FHR averaged over the full recording). Fetal movement data represent raw voltage values generated from the actograph calibrated by multiplying by a conversion factor and scaled from 0 to 100 in arbitrary units (a.u.s). Fetal motor activity was computed as the mean signal values generated per minute, averaged over the entire recording. The relation between FHR and motor activity (hereafter, FM–FHR coupling), was calculated as the proportion of discrete FM bouts, defined as commencing when an actograph signal attained an amplitude of 15 units and ending with a cessation of signal for at least 10 s, that were associated with excursions in FHR ≥ 5 bpm over baseline within 5 s before the start of a movement or within 15 s after the start of a movement, consistent with previously developed criteria (Baser, Johnson, & Paine, 1992; DiPietro et al., 1996).

Postnatal. Infants were examined using the Dubowitz Neurological Examination of the Newborn (Dubowitz et al., 1998) within the first 2 weeks after birth during a home visit conducted by a pediatric nurse practitioner. The assessment is a more intensive version of exams routinely used to assess status and gestational age immediately postpartum. It contains 34 items distributed into six clusters: Tone (e.g., posture and head control), Tone Patterns (e.g., extensor vs. flexor), Reflexes (e.g., Moro, placing), Motor (e.g., spontaneous movements and head raising), Abnormal Signs (e.g., tremoring), and Behavior (e.g., orientation, irritability). Scoring proceeded based on optimality criteria, developed by validation across gestational age (Dubowitz et al., 1998). That is, raw scores are coded along an optimality continuum such that a higher score on an item is not necessarily coded as more optimal.

A second visit was scheduled to collect BAEP. Initially, the protocol was designed to accomplish both at one visit; however, this interfered with obtaining BAEP data of sufficient quality. A licensed audiologist was subsequently dispatched to conduct a second visit but this change was not implemented until the 21st participant. BAEP measurements were undertaken utilizing single-channel differential recording with an active electrode attached to the high forehead and a reference electrode attached to the mastoid. The contralateral electrode served as ground. Testing relied on a portable BAEP unit (Smart USB Lite; Intelligent Hearing Systems, Miami, FL) connected to a laptop computer through a USB port to a single channel amplifier. Infants were generally swaddled and placed in a supine position or in their mothers' arms. Recording took place while infants were in a sleep state. Stimuli were rarefaction clicks presented at a rate of 27.7/s through ER3 insert phones at a level of 70 dB nHL. Responses to two trials of 2,048 clicks were filtered (30–1500 Hz) and amplified (100 K). A 12-ms recording window was utilized. The analysis was undertaken offline (SmartEP System, version 3.82; Intelligent Hearing Systems) and latencies for Waves I, III, and V were assessed for each trial and each ear. Both trials were averaged. No significant differences in values generated between right and left ears were detected, so these were further averaged. Interpeak latency intervals were computed by subtracting the earlier wave from the later. The following variables were used in the analysis: Wave V latency (i.e., stimulus presentation to Wave V peak) and interpeak intervals I–III, III–V, and I–V.

Analysis Plan

Variables were examined for skewness and outliers. Hierarchical linear models (HLMs) were estimated using SPSS Mixed (version 15.0) to examine growth in fetal neurobehavioral development across the second half of gestation. Models incorporated two levels: Level 1 models the change in the dependent variable over time for each individual in the sample, whereas Level 2 relates individual growth parameters to predictors of interest (Bryk & Raudenbush, 1987; Singer & Willett, 2003). Two advantages emerge from this approach. First, individuals are not directly compared based on the dependent variables of interest but on the parameters describing their developmental slope across gestation. Therefore, week of gestation at observation was allowed to vary across participants and is accounted for in the Level 1 model. Second, the number of data points per subject can also vary as long as there is sufficient information to describe the developmental trajectory for that individual. Unbalanced data sets, including those with planned missing data as generated by the current design, are thus acceptable in these models (Singer & Willett, 2003).

Restricted maximum likelihood was used in reporting model parameters when assessing the significance of the random effects; degrees of freedom were estimated using the Satterthwaite method. The 95% confidence interval (CI) for random variation around each fixed effect was calculated as ± 2 *SD* of its accompanying random variance term. To best describe growth trajectories, linear, polynomial, and piecewise (spline) terms were examined during model fitting. Models were centered at 38 weeks, such that the intercept represented individual differences in fetal neurobehaviors (FHR variability, motor activity, and FM–FHR coupling) at the end of gestation. Models were initially specified with random intercepts, and the effects of including random slopes on model fit were subsequently evaluated. The fixed effects of each predictor (e.g., fetal sex, maternal distress) on the intercept and change across gestation, calculated as slope from 24 to 38 weeks, were tested sequentially in the whole group. Model fit was assessed with likelihood deviance difference tests for nested models (Singer & Willett, 2003). This process evaluates the significance of the change in model fit following the addition of a new parameter(s) (e.g., a quadratic term) to a previously tested model (e.g., a linear model).

Exploratory analyses were used to identify potential covariates. No outliers in postnatal measures were detected. Random intercepts and slopes exported from the HLM models described above were used in correlation and regression analyses.

Pearson correlations were computed between fetal and infant measures. Regression models were constructed by entering covariates at the initial step (e.g., gestational age) followed by entry of the fetal predictor (i.e., intercept or slope for FHR variability, FM–FHR coupling, and fetal motor activity).

Results

Fetal Neurobehavioral Development From 24 to 38 Weeks Gestation

Nearly half (49.1%) of the participants completed all five data visits, 38.4% completed four visits, and 12.5% completed three visits. Of those with missed visits, 13 (22.8%) were a result of delivery of infant prior to the last (i.e., 36–38 weeks) scheduled visit. Scheduling problems associated with the assessment schedule of 3-week intervals were the most common reason for the remaining missed visits. HLMs were fit to produce estimates of individual differences in fetal neurobehaviors at 38 weeks (intercept) and growth across the second half of gestation. Model parameters are given in Table 1. Unadjusted raw means, best fit models, and CIs based on individual values are presented in Figure 1. No sex differences in intercept or developmental slope were detected for fetal measures.

Fetal heart rate variability. Fetal heart rate variability increased linearly from 24 to 38 weeks gestation ($p < .001$) by 0.27 bpm (4.5%) per week (95% CI: 0.22–0.31) without any significant changes in slope (Figure 1a), to reach an average of 9.95 bpm by 38 weeks (95% CI: 9.44–10.47).

Fetal motor activity. Ten visits were excluded for fetal motor activity due to technical difficulties with the FM signal and an additional motor activity outlier (+3 *SD*) was removed at 33 weeks. The amplitude of fetal motor activity increased gradually at a rate of 0.04 a.u.s. (0.7%) per week (95% CI: 0.01–0.07) with advancing gestation ($p < .05$; Figure 1b). By 38 weeks, mean fetal motor activity reached a level of 5.66 a.u.s. (95% CI: 5.37–5.95 a.u.s.).

FM–FHR coupling. A single outlier (+3 *SD*) was excluded from the FM–FHR coupling analysis at 34 weeks. FM–FHR coupling increased steadily from 24 to 32 weeks gestation ($p < .001$), with a significant decline in slope after 32 weeks (32-week spline term, $p < .01$; Figure 1c). The coupling index initially increased at a rate of 0.014 (8.0%) per week of gestation (95% CI: 0.011–0.017). After 32 weeks, the rate of increase declined by 0.010 per week (95% CI: −0.003 to −0.017) to 29% of the original slope. By 38 weeks, the mean FM–FHR coupling index had reached 0.309 (95% CI: 0.287–0.330).

Random linear slopes improved the fit for FHR variability, motor activity, and FM–FHR coupling models (*p*s < .05). Additional piecewise models were generated for FM–FHR coupling to obtain individual differences in linear slopes before and after 32 weeks.

Postnatal Assessment

Infant data, stratified by sex, are presented in Table 2. Infants were normally grown and none were low birth weight. There were four instances of mildly preterm delivery (two each at 35 and 36 weeks gestation); all were discharged from the

Table 1
Hierarchical Linear Models Describing Change in Fetal Heart Rate Variability, Fetal Motor Activity, and FM–FHR Coupling Across the Second Half of Gestation

	Fetal heart rate variability			Fetal motor activity			FM–FHR coupling		
Fixed effects	Estimate	*SE*	*T*	Estimate	*SE*	*t*	Estimate	*SE*	*t*
Intercept at 38 weeks	9.95	0.26	38.48***	5.66	0.15	38.40***	0.308	0.011	28.19***
Linear slope	0.27	0.02	11.93***	0.04	0.02	2.43*	0.014	0.002	8.25***
Slope deviation at 32 weeks							−0.010	0.004	−2.90**
Random effects	Estimate	*SE*	Wald Z	Estimate	*SE*	Wald Z	Estimate	*SE*	Wald Z
Residual	3.00	0.25	11.81***	1.34	0.12	11.38***	0.006	0.000	11.74***
Intercept variance	4.61	1.03	4.49***	1.06	0.33	3.23**	0.005	0.001	3.39*
Linear slope variance	0.02	0.01	1.97*	0.01	0.00	2.28*	0.000	0.000	1.13
Intercept–linear slope covariance	0.26	0.08	3.16**	0.07	0.03	2.28*	0.000	0.000	2.15*

Note. FM = fetal movement; FHR = fetal heart rate.
*$p < .05$. **$p < .01$. ***$p < .001$.

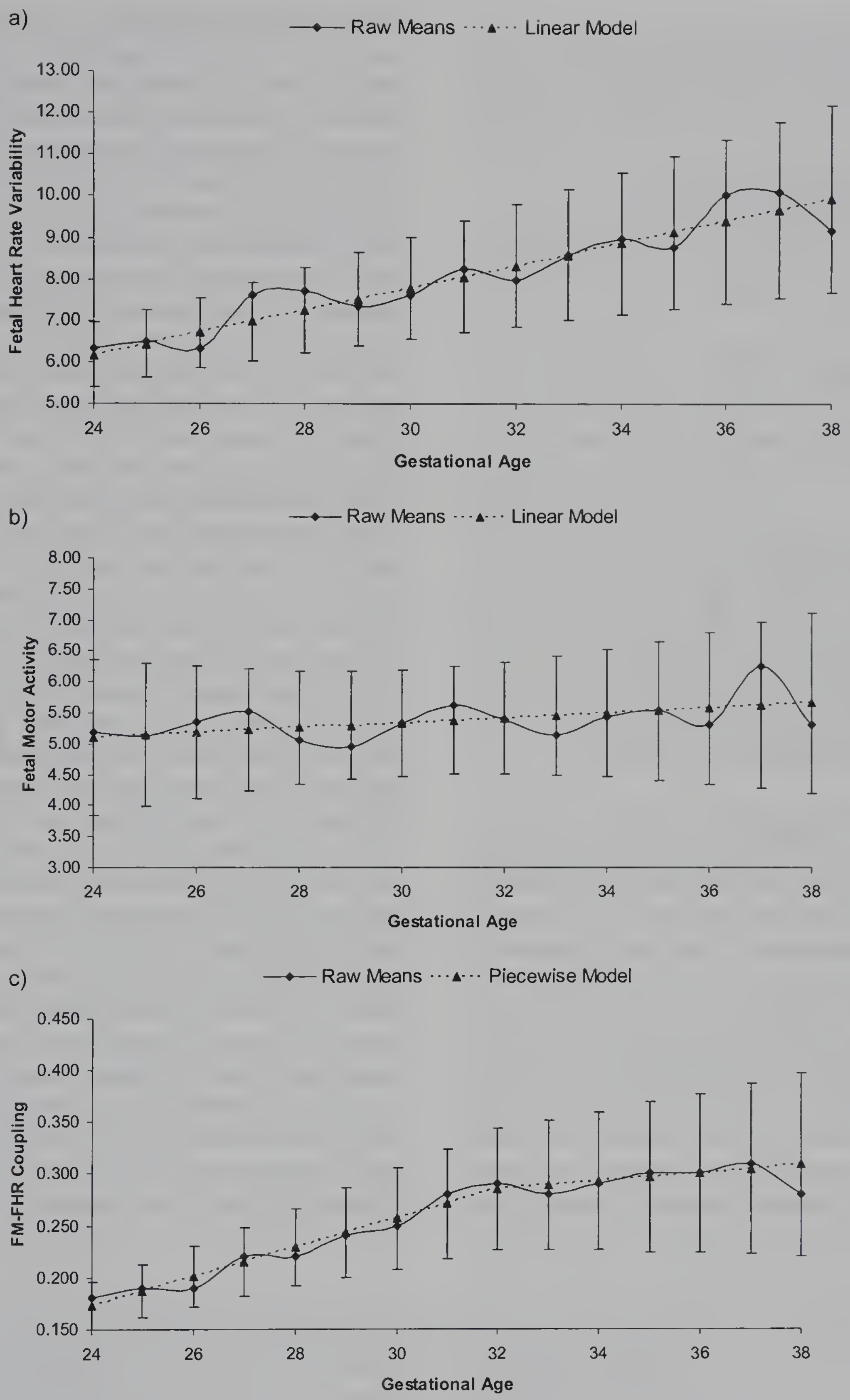


Figure 1. Raw means and best fit models with 95% confidence intervals based on individual values for fetal neurobehaviors across the second half of gestation. Fetal heart rate variability (a) and motor activity (b) increased linearly from 24 to 38 weeks gestation. Fetal heart rate–movement coupling (c) increased from 24 to 32 weeks, followed by a decline in rate after 32 weeks.

hospital on routine schedules. Twenty-four percent (n = 27) were delivered by Caesarian section. Six infants were determined to be too sick to visit and test (e.g., thrombocytopenia, sepsis) during the evaluation period. There were no sex differences on birth outcome measures.

Dubowitz Neurological Exam. Ninety-seven of the remaining 106 newborns received the Dubowitz Neurological Exam during the home visit (91.5%). Reasons for lack of visiting included nurse unavailability (n = 5), residence relocation out of local area (n = 2), or declining of evaluation (n = 2). Infants

Table 2
Descriptive Birth and Neonatal Assessment Values

			Boys		Girls	
	M	*SD*	*M*	*SD*	*M*	*SD*
Birth characteristics (n = 112; 55 boys, 57 girls)						
Gestational age (weeks)	39.4	1.2	39.4	1.4	39.5	1.1
Birth weight (g)	3,443.5	435.6	3,439.0	468.3	3,447.9	405.8
Head circumference (cm)	34.2	1.5	34.1	1.5	34.3	1.7
5-min Apgar	8.9	0.4	8.9	0.5	8.9	0.3
Dubowitz optimality scores (n = 97; 47 boys, 50 girls)						
Tone	8.6	1.5	8.4	1.7	8.9	1.2
Tone Patterns	4.4	0.7	4.3	0.8	4.5	0.6
Reflexes	5.3	0.6	5.3	0.7	5.2	0.5
Behavior	5.6	1.4	5.5	1.4	5.6	1.3
Motor	2.2	1.0	2.0	1.1	2.3	1.0
Brainstem auditory evoked potential (n = 47; 24 boys, 23 girls)						
Axillary temperature	97.7	0.7	97.6	0.5	97.7	0.8
Wave V	7.3	0.3	7.4	0.3	7.1	0.3[a]
Interpeak interval I–V	5.2	0.3	5.3	0.3	5.0	0.3[a]
Interpeak interval I–III	2.8	0.2	2.9	0.2	2.7	0.2[a]
Interpeak interval III–V	2.4	0.2	2.4	0.2	2.4	0.2

[a] Signifies significant sex difference; see text for details.

were examined between Days 2 and 9 postpartum ($M = 5.0$, $SD = 1.2$); 78% were tested on Day 4, 5, or 6. There were no significant, or near significant correlations between age at test and any optimality cluster, rs(95) range from −.10 to .06. One of the six optimality cluster scores (i.e., Abnormal Signs) did not have sufficient variability to analyze so was discarded. Remaining cluster values are presented in Table 2; there were no sex differences.

Fetal motor activity level was positively associated with optimality in the Motor and Reflex clusters, $r = .27$, $p < .01$, and $r = .20$, $p < .05$, respectively. Exploratory analysis revealed that a number of birth characteristics were also associated with some of the exam clusters, including gestational age and delivery type (vaginal vs. Caesarian). In general, infants with longer gestations scored more optimally whereas those delivered by Caesarian section scored less optimally. Regression results, controlling for gestational age, delivery type, and infant sex are presented in Table 3. In general, more active fetuses displayed more optimal motor and reflex patterns as neonates. In addition to level, analysis of associations between the slope of fetal motor activity during gestation indicated positive associations with Motor cluster optimality, $r(95) = .20$, $p < .05$, but the $r^2\Delta$ did not remain significant in the regression model, $F(4, 92) = 3.04$, $p < .10$. However, a significant negative association between Tone Pattern optimality and fetal activity slope was also detected, $r(95) = -.23$, $p < .05$, and contributed additional significant variance to the regression model, multiple $R = .47$, $r^2\Delta = .06$, $F(4, 92) = 6.57$, $p < .01$; there was no association between Tone Pattern and fetal activity level.

The fetal motor measure used in the modeling that generated these gestational values is an indicator of overall motor signal output and does not capture all characteristics of fetal motor behavior. Supplemental correlations were computed between four additional measures of motor behavior generated at the last fetal assessment and the Motor and Reflex optimality scores to determine the generalizability of the motor results. Fetal motor measures were contributed by 86 fetuses at this assessment, of those, 77 had a postnatal exam. Significant associations were detected for three of these measures with postnatal Motor optimality: mean duration of individual movements, the longest movement bout during the recording, and the total time, in minutes, spent moving, rs(75) = .25, ps < .05. The number of movement bouts was unrelated to Motor optimality scores, and no cross-sectional motor characteristics

Table 3
Multiple Regression Models Predicting Newborn Motor and Reflex Optimality From Fetal Motor Activity Level (N = 97)

	Motor optimality				Reflex optimality			
	β	*t*	*p*	Multiple *R*	β	*t*	*p*	Multiple *R*
Constant	−9.68				1.63			
Infant sex	0.25	1.26	*ns*		−0.16	−1.43	*ns*	
Delivery type	−0.65	−2.75	< .01		−0.12	−.88	*ns*	
Gestational age	0.25	3.06	< .01	.37	0.07	1.55	*ns*	.20
	$R^2 = .13$, $F(3, 93) = 4.83$**				$R^2 = .04$, $F(3, 93) = 1.35$			
Fetal motor activity	0.39	2.88	< .01	.45	0.16	2.07	< .05	.29
	$R^2\Delta = .07$, $F(4, 92) = 8.29$**				$R^2\Delta = .04$, $F(4, 92) = 4.28$*			

*$p < .05$. **$p < .01$.

at term, other than the longitudinally modeled measure, were related to Reflex optimality.

Significant, negative associations were detected between the level and developmental slope of FHR variability with Tone Optimality, $rs(95) = -.24$, $ps < .05$. Both level, multiple $R = .32$, $R^2\Delta = .047$, $F(4, 92) = 4.85$, $p < .05$, and slope, multiple $R = .31$, $R^2\Delta = .045$, $F(4, 92) = 4.63$, $p < .05$, contributed significant unique variance to the regressions. FM–FHR coupling intercept and slope were unrelated to Optimality scores.

Brainstem auditory evoked potential. As described in the Procedures subsection, dedicated BAEP visits commenced partway into the protocol. Testing occurred between postpartum Days 3 and 14 ($M = 8.6$, $SD = 2.8$; M interim between first and second visits = 3.5 days, $SD = 2.7$). There were 70 visits in which BAEP data collection was attempted. One parent declined the procedure, and poor signal quality or recording problems (e.g., persistent 60 cycle noise, lack of infant sleep state, motor artifact, or electrical interference from household sources that could not be determined) resulted in inability to collect complete data (i.e., two trials of 2,048 stimuli presentations per ear) at 22 of the visits, resulting in 47 infants with BAEP data. Axillary temperature was recorded for each infant at the start of the visit because of an existing known relation between body temperature and BAEP conductance (Bastuji, Larrea, Bertrand, & Mauguiere, 1988; Litscher, 1995). Infants on whom BAEP data could and could not be collected did not differ on birth characteristics (e.g., birth weight, gestational age, and delivery method), sex, or Dubowitz cluster scores. The only significant difference was postpartum age at which the data collection was attempted: Successful recordings were more likely to be generated on slightly older infants, $M = 9.4$ versus 7.6 days; $t(65) = 2.67$, $p < .01$. Mean values are presented in Table 2.

Gestational age, postnatal age at testing, and measures of size taken at the prior visit were unrelated to BAEP waves, with one modest exception. Greater head circumference was negatively associated with interpeak latency III–V, $r(45) = -.32$, $p < .05$. However, boys showed significantly slower BAEP latencies for Wave V, $t(45) = 3.03$, $p < .01$, and interpeak intervals I–V, $t(45) = 2.93$, $p < .01$ and I–III, $t(45) = 2.81$, $p < .01$. Infants with higher temperatures tended to have longer latencies, rs ranged from .21 to .32 for Wave V and interpeak latencies I–III and I–V, although only the last value attained significance ($p < .05$). Table 4 presents partial correlation coefficients, controlling for temperature, between FHR variability, and FM–FHR coupling intercepts and slopes for each BAEP measure. In general, higher levels and steeper developmental trajectories for FHR variability and FM–FHR coupling were predictive of shorter interpeak intervals from Waves I and III to Wave V. Fetal motor activity was unrelated to BAEP.

Table 4

Partial Correlations Between Fetal Neurobehavioral Measures and BAEP Values Controlling for Infant Temperature (N = 47)

		Interpeak latencies		
	Wave V	I–III	III–V	I–V
FHR variability				
Level	–.36*	–.13	–.30*	–.32*
Slope	–.36*	–.15	–.29*	–.32*
FM–FHR coupling				
Level	–.29*	.00	–.42**	–.28$^+$
Slope < 32 weeks	–.11	–.07	–.34*	–.12
Slope ≥ 32 weeks	–.36*	.14	–.42**	–.34*

Note. BAEP = brainstem auditory evoked potential; FHR = fetal heart rate; FM = fetal movement.
$^+p < .10$. $^*p < .05$. $^{**}p < .01$.

Correspondence Between Dubowitz Exam and BAEP Assessment

No association between any cluster score of the Dubowitz neurologic exam and any BAEP variable was detected.

Associations Between Maternal Psychological Distress and Development

Fetal neurobehavioral development. HLMs did not reveal any change over gestation for either pregnancy-specific hassles (estimate = 0.00, $SE = 0.00$, $t = 1.58$, *ns*) or trait anxiety (estimate = 0.02, $SE = 0.04$, $t = 0.61$, *ns*). Pregnancy-specific uplifts gradually increased (estimate = 0.01, $SE = 0.00$, $t = 3.11$, $p < .01$). Variables were centered at the grand mean, and average scores were used in analysis of associations with fetal neurobehavioral measures. Means are as follows: PES intensity of hassles, $M = 1.41$, $SD = 0.39$; PES intensity of uplifts, $M = 2.37$, $SD = 0.41$; and trait anxiety, $M = 31.52$, $SD = 7.13$.

Trait anxiety and pregnancy uplifts were not associated with level or change for any fetal neurobehavior. Because chest wall motions generated by fetal hiccups can generate significant artifact in the motor signal, the presence of fetal hiccups in 59 of the visits was controlled for by inclusion of a dummy variable in these analyses. A greater intensity of pregnancy hassles (1 *SD* above the mean) was predictive of 2.8% higher FHR variability (95% CI: 0.2–5.4), $t = 2.17$, $p < .05$, and marginally higher motor activity (2.5%; 95% CI: −0.0 to 5.2), $t = 1.86$, $p = .065$ in models centered at 38 weeks (Figures 2a and 2b). These associations were confirmed by correlations between these two fetal measures and the mean score for intensity of hassles, $rs(110) = .19$, $ps < .05$. In addition, more intense pregnancy hassles (+1 *SD* above the mean) were also associated with an increase in the slope of FM–FHR coupling after 32 weeks by 0.010 per + 1*SD* per week (95% CI: 0.001–0.019), $t = 2.10$, $p < .05$, augmenting it to twice the post-32-week slope. Conversely, less intense pregnancy hassles (−1 *SD* below the mean) were associated with a decline of 0.010 per 1 *SD* per week, resulting in a nearly flat FM–FHR coupling slope after 32 weeks (i.e., 0.0001 increase per week).

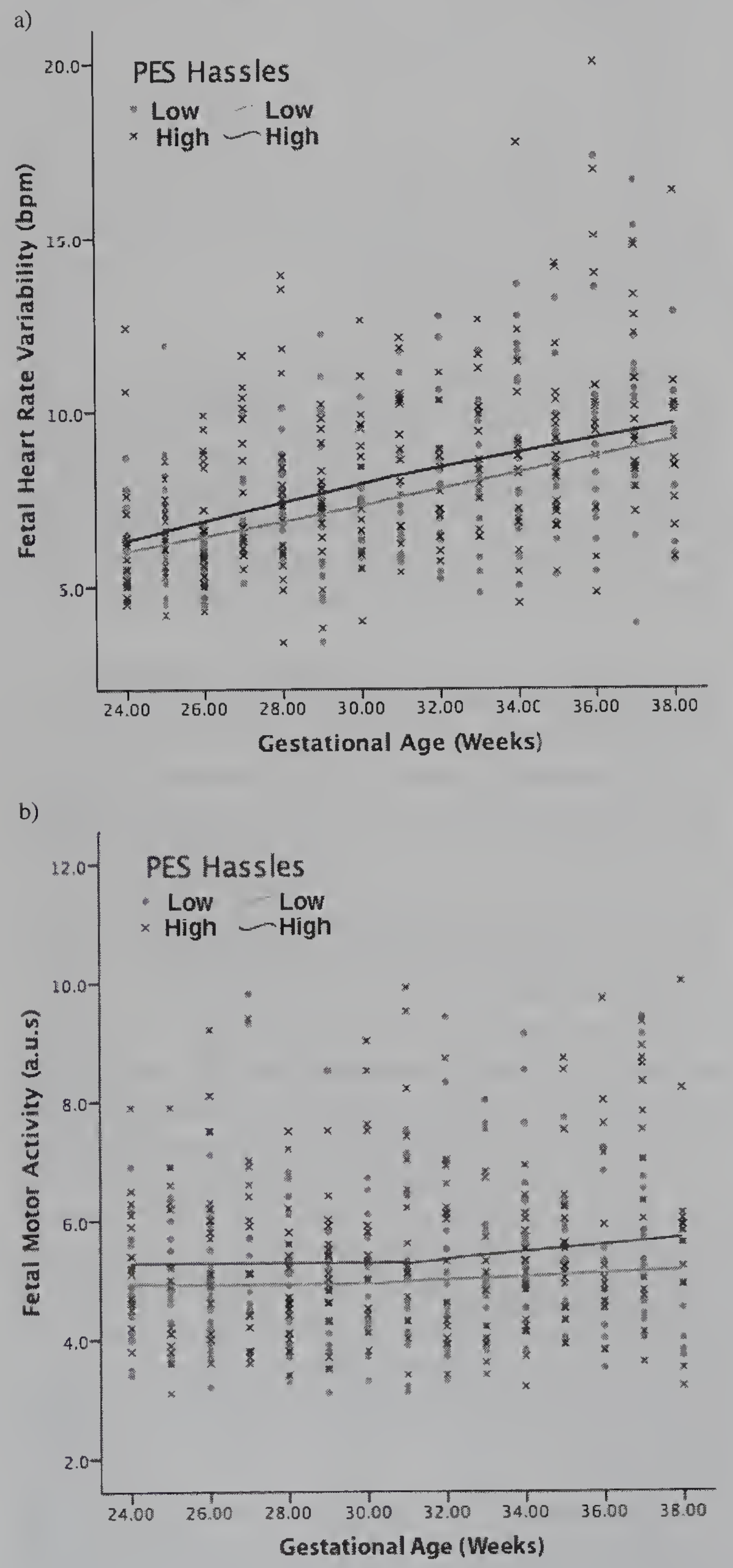


Figure 2. Fetuses of mothers reporting a greater intensity of pregnancy-specific hassles (1 *SD* above the mean; solid black line) show significantly higher heart rate variability (a) and more fetal motor activity (b) during the second half of gestation.

Postnatal maturation. PES–Brief hassles, uplifts, and STAI anxiety scores were unrelated to the motor and tone-related optimality clusters of the Dubowitz exam. However, exploratory data analyses indicated significant associations between PES–Brief hassles intensity and three components of the Behavior optimality cluster: irritability, consolability, and crying, but not with items related to alertness or orientation performance. A composite score for irritability was derived from these three items; the association between it and PES hassles was significant, $r(95) = .28$, $p < .01$. Separate analysis by sex revealed that this association was significant for boys, $r(45) = .45$, $p < .01$, but not for girls, $r(48) = .08$. The coefficients were significantly different, $Z = 1.75$, $p < .05$, one-tailed.

There was a significant pattern of associations between PES–Brief hassle intensity scores and BAEP values as follows: Wave V, $r(45) = -.35$, $p < .05$; interpeak latencies I–III, $r(45) = -.27$, $p < .10$; and interpeak latencies I–V, $r(45) = -.30$, $p < .05$. There were no significant associations for PES–Brief uplifts or STAI anxiety scores. Analyses conducted separately by sex indicated augmentation of these associations for boys, $rs(22)$ ranged from −.35 to −.45. In contrast, none of the associations with PES hassles was significant for girls. However, significant associations emerged for anxiety scores, Wave V and interpeak intervals I–V, III–V, $rs(21)$ ranged from −.48 to −.59.

Discussion

The current study provides confirmation of the three working hypotheses and underscores the role of the fetal period in providing the foundation for postnatal development. The observed patterns of associations were consistent with the prediction that fetal motor activity would be associated with newborn maturation as assessed by a traditional examination but that FHR variability and somatic–cardiac coupling would differentially predict BAEP. With respect to the first of these, the current results support the long-standing speculation that prenatal motor activity is an ontogenic adaptation that prepares the fetus for postnatal life (Prechtl, 1984). There are at least two interpretations of the observed associations. The first is that the relation between greater fetal motor activity and more optimal neonatal motor and reflex performance simply reveals conservation in motor functioning from the prenatal to postnatal periods and reflects an intrinsic characteristic of the individual perhaps as a result of a genetic or other constitutional contribution. The second possibility is that greater fetal motor activity provides a practice effect that develops the musculature and potentiates the neural circuitry associated with more optimal neonatal motor performance and reflex responsivity. Based on psychobiological observations of the epigenetic determination of prenatal motor behavior in animal preparations (Smotherman & Robinson, 1987), it is likely that both interpretations contribute to the observed findings. There was an unexpected negative association between steeper developmental slopes for motor activity and lower newborn Tone Pattern. Examination of the scale scores indicates that lower Tone Pattern scores were generally reflective of hyperresponsivity in flexor or extensor tone. Given that there was no association with fetal motor activity level, this suggests that an unusually steep developmental trajectory of motor activity may reflect an inoptimal developmental progression that results in excess tone. Similarly, we have no ready explanation for the negative associations between FHR variability, a measure that is typically regarded as an indicator of autonomic integration, and Tone optimality.

The second set of findings relates measures of fetal neurobehavioral functioning that have been conceptually linked to maturation of the nervous system with measurement of neural conductivity. Higher levels and steeper slopes of both FHR variability and FM–FHR coupling during development were associated with shorter neural conduction times for Wave V latency in general as well as the Wave III–V interpeak interval. The intervals between the component waves provide an estimate of the rate at which signals progress from peripheral to more central levels of the auditory pathway. Given their anatomical origins, and parallels with the progression of development throughout the nervous system, latencies of the later components of the BAEP (i.e., Waves III and V) are the least mature in newborn infants. This reflects incomplete neural myelination of the fibers in the auditory pathway, reduced axon diameter, and immaturity in synaptic function (Eggermont & Salamy, 1988; Folsom & Wynne, 1987) that mature over the first 18 months of life.

The associations detected in the current study between fetal measures and conduction of Wave V and the III–V interval reflect activity focused in the later maturing regions of the brainstem auditory pathway. The relative degree of immaturity in these components in the newborn ensures greater detection of interindividual variation. The notion that BAEP provides a marker of functional integrity of the nervous system beyond the confines of the auditory pathway has been supported by studies that find reductions in auditory pathway conduction following prenatal exposures known to negatively impact postnatal development, such as lead (Rothenberg et al., 2000) and accelerated conductivity following postnatal exposures that enhance development, such as supplementation with docosahexaenoic acid (Unay et al., 2004). BAEP also corresponds to neurodevelopmental pathologies (Jiang et al., 2008). With respect to the current findings, it should be noted that the auditory brainstem pathway lies in close proximity to neural units that control heart rate and breathing (Amin, Charafeddine, & Guillet, 2005). More importantly, the predicted absence of any associations with fetal motor activity and the consistent associations with both FM–FHR coupling and FHR variability confirm speculation that the latter measures better reflect neural integrity of the developing fetus whereas fetal motor activity may more accurately provide an indication of temperament. However, this discussion needs to be qualified by the fact that less than half the original sample generated BAEP data so findings should be regarded as more preliminary than definitive. Home recording of evoked potentials with portable equipment is a particularly challenging undertaking as BAEPs consist of bioelectrical events in the microvolt range and are highly susceptible to interference from ambient and electrical noise in the home environment. Nonetheless, the fairly consistent findings provide support for conservation and/or prediction of neonatal neural maturation from fetal indicators.

The prenatal models generated from this study also serve to extend our prior work documenting normal ontogeny of fetal neurobehavioral development by providing continuously generated gestational data in weekly intervals. In general, these confirmed developmental trajectories that had previously been generated by data collected in monthly intervals. These include significant increases in FHR variability and FM–FHR coupling and little or no change in fetal motor activity (DiPietro et al., 2004). There was one exception; in prior studies, significant changes in slope were detected at 32 and 28 weeks gestation for FM–FHR coupling and FHR variability, respectively, such that the rate of development slowed near the end of gestation. The current study confirmed the developmental transition for FM–FHR coupling but not variability. This suggests that the longer interval between sampling in the prior report, which was 4 weeks, may have contributed to the perception of discontinuity in contrast to the more gradual change revealed by weekly measurement.

The third aim of the current study was to investigate the contributory role of maternal stress and anxiety on both the fetus and newborn. Stress is a broad construct to operationalize; in this study we focused on the types of stressors, both positive and negative, that are specific to pregnancy, as these have particular salience to the pregnant woman. For the sample as a whole, significant associations were detected with maternal stress but not anxiety. Fetuses of women with higher levels of maternal pregnancy-specific stress displayed higher levels of FHR variability and steeper incline in somatic–cardiac coupling as term approached. Variability in heart rate has been well documented as an indicator of parasympathetic maturation in both the prenatal (Martin, 1978) and postnatal (Bar-Haim, Marshall, & Fox, 2000; Porges, 1983) periods. Increases in synchronous activation of somatic and cardiac processes is a conspicuous feature of neurobehavioral development during gestation and has been similarly implicated as a marker for neural integration (Johnson et al., 1992) and fetal well-being (Baser et al., 1992). Thus, contemporaneous maternal stress appears to have facilitative effects on fetal neurodevelopment.

In addition, fetuses of women with higher stress were found to be marginally more active; replicating an earlier report using the same measure of maternal stress (DiPietro, Hilton, Hawkins, Costigan, & Pressman, 2002). Unlike the other two fetal measures, there is no compelling conceptual basis to regard variation in fetal motor activity along a continuum of maturation such that higher activity is either better or worse than lower activity. Instead, as in infants, fetal motor activity may best be considered as an indicator of temperament (Eaton & Saudino, 1992). Two studies have reported that maternal cortisol is positively correlated with fetal activity level (DiPietro, Kivlighan, Costigan, & Laudenslager, 2009; Field, Diego, Hernandez-Reif, Gil, & Vera, 2005), suggesting a potential mechanism for the effects observed here.

In the postnatal period, maternal prenatal stress was positively associated with neonatal irritability. This finding partially supports others that detect associations between higher maternal prenatal stress or anxiety and less infant behavioral regulation and/or greater reactivity to novelty measured by observational methods (Davis et al., 2004; Gutteling et al., 2005; Huizink et al., 2002) although another study reports the reverse association (Mohler, Parzer, Brunner, Wiebel, & Resch, 2006). However, in the current study, prenatal stress was also associated with faster neural conduction for three of the four BAEP latencies. In a prior report on a different sample, we found that maternal prenatal stress was associated with higher Bayley Scales of Infant Development scores at age 2, after controlling for postnatal maternal psychological measures (DiPietro et al., 2006). However, there is significant concern regarding the degree to which indicators of mental and motor development as measured by standard developmental assessments adequately reflect information processing, which is at the core of early brain development. Thus, BAEP results in the present study provide perhaps more compelling support for the maturative role of maternal prenatal stress on development as BAEP was measured closer to birth and is unambiguously representative of neural processing of information. Together, the postnatal findings suggest that dimensions of information processing and temperament should be regarded as fairly distinct systems, and efforts to measure the former should be careful to exclude confounding influences of the latter.

The current literature detailing maternal stress effects on human development is long on theories of why maternal stress is damaging to the fetus but short on empirical data. Most conceptualization is based on animal models, which use an experimental approach to delivering repetitive, stressful events. These are quite different from human models, which necessarily rely on an observational approach to maternal appraisal of elements within their lives, which in turn is confounded by dispositional features of maternal temperament (DiPietro, 2004). Although evidence from animal models often supports deleterious consequences in experimental paradigms (e.g., Weinstock, 2001), functional deficits are not uniformly found. For example, rodents subjected to mild stress showed better spatial learning coupled with facilitated differentiation in neuronal morphology (Fujioka et al., 2001) and more exploratory behavior (Meek, Burda, & Paster, 2000). Moreover, methodologies routinely used in human studies have significant limitations. Some of the most widely cited studies in support of an adverse role of maternal stress on behavioral or cognitive development include the following design problems: reliance on maternal report of child outcomes, as opposed to observational data collection, thereby inextricably confounding the dependent and independent (i.e., maternal psychological distress) measures in a direction that promotes detection of positive associations; failure to adequately control for postnatal maternal psychological distress, which confers well-known, independent effects on child development as a result of environmental, as opposed to biological, mediation; and lack of control for disparities in social class within samples when it is linked to both exposure and outcomes.

Although the findings reported herein may be counterintuitive to some readers, they are consistent with the orientation that the human brain requires sufficient, but not excessive, stress to promote neural development both before (Amiel-Tison et al., 2004) and after (Huether, 1998) birth. It is important to note that the sample in this study is composed of well-educated, financially stable women with well-nourished pregnancies, and as such, the findings may not generalize to populations of women with psychopathology or to those that experience the types of chronic stressors associated with poverty. In terms of a recently proposed lexicon for the consequences of stress exposure on child development, fetal exposures to stress in this sample would likely fall into the categories of "positive" (i.e., moderate, short-lived exposures) or "tolerable" (i.e., time limited and within a supportive environment) as opposed to "toxic" levels that exceed the capabilities of full recovery (Shonkoff, 2006).

No sex differences in development during the fetal period were evident, confirming findings reported on other samples of similar size (DiPietro et al., 2004; Robles de Medina, Visser, Huizink, Buitelaar, & Mulder, 2003). In the postnatal period, however, boys had significantly slower BAEP conduction in three of the four latency measures, suggesting slower neural maturation. Given the equivalence in anthropometric indicators of growth between boys and girls, these findings cannot be attributed to conductance differences secondary to size. Instead, they support the long-standing recognition that girls are more mature than boys at birth (Tanner, 1978). In addition, there was differential effect of pregnancy-specific stress exposure on both types of newborn outcomes such that the positive associations between stress and neonatal irritability and neural conduction scores were greater for boys than girls. A pattern of differential responsiveness in neonatal maturity to maternal stress hormones between boys and girls has recently been shown (Ellman et al., 2008). These results suggest that features of the intrauterine environment may interact with fetal sex in ways that remain largely unknown and unexplored.

In summary, the current results provide normative data regarding predictors of neonatal maturation in a sample of healthy women with normally progressing pregnancies. This information provides a basis for future examination of the manner in which early experiences associated with adversity in the maternal environment, intrauterine stress, or exposure to potentially deleterious substances may exert influence on the fetus, and in turn, the child.

References

Alder, J., Fink, N., Bitzer, J., Hosli, I., & Holzgreve, W. (2007). Depression and anxiety during pregnancy: A risk factor for obstetric, fetal and neonatal outcome? A critical review of the literature. *Journal of Maternal-Fetal & Neonatal Medicine*, *20*, 189–209.

Almli, C. R., Ball, R. H., & Wheeler, M. E. (2001). Human fetal and neonatal movement patterns: Gender differences and fetal-to-neonatal continuity. *Developmental Psychobiology*, *38*, 252–273.

Als, H. (1982). Toward a synactive theory of development: Promise for the assessment and support of infant individuality. *Infant Mental Health Journal*, *3*, 229–243.

Amiel-Tison, C., Cabrol, D., Denver, R., Jarreau, P., Papiernik, E., & Piazza, P. (2004). Fetal adaptation to stress part I: Acceleration of fetal maturation and earlier birth triggered by placental insufficiency in humans. *Early Human Development*, *78*, 15–27.

Amiel-Tison, C., Gosselin, J., & Kurjak, A. (2006). Neurosonography in the second half of fetal life: A neonatologist's point of view. *Journal of Perinatal Medicine*, *34*, 437–446.

Amin, S. B., Charafeddine, L., & Guillet, R. (2005). Transient bilirubin encephalopathy and apnea of prematurity in 28 to 32 weeks gestational age infants. *Journal of Perinatology*, *25*, 386–390.

Amin, S. B., Orlando, M. S., Dalzell, L. E., Merle, K. S., & Guillet, R. (1999). Morphological changes in serial auditory brain stem responses in 24 to 32 weeks' gestational age infants during the first week of life. *Ear and Hearing*, *20*, 410–418.

Bar-Haim, Y., Marshall, P. J., & Fox, N. A. (2000). Developmental changes in heart period and high frequency heart period variability from 4 months to 4 years of age. *Developmental Psychobiology, 37*, 44–56.

Barker, D. J. P. (2006). Adult consequences of fetal growth restriction. *Clinical Obstetrics and Gynecology, 49*, 270–283.

Baser, I., Johnson, T. R. B., & Paine, L. L. (1992). Coupling of fetal movement and fetal heart rate accelerations as an indicator of fetal health. *Obstetrics and Gynecology, 80*, 62–66.

Bastuji, H., Larrea, L., Bertrand, O., & Mauguiere, F. (1988). BAEP latency changes during nocturnal sleep are not correlated with sleep stages but with body temperature variation. *Electroencephalography and Clinical Neurophysiology, 70*, 9–15.

Besinger, R. E., & Johnson, T. R. B. (1989). Doppler recordings of fetal movement: Clinical correlation with real-time ultrasound. *Obstetrics and Gynecology, 74*, 277–280.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101*, 147–158.

Buitelaar, J., Huizink, A., Mulder, E., Robles de Medina, P., & Visser, G. (2003). Prenatal stress and cognitive development and temperament in infants. *Neurobiology of Aging, 24*, S53–S60.

Davis, E., Snidman, N., Wadhwa, P., Glynn, L., Dunkel-Schetter, C., & Sandman, C. (2004). Prenatal maternal anxiety and depression predict negative behavioral reactivity in infancy. *Infancy, 6*, 319–331.

DiPietro, J. A. (2004). The role of prenatal maternal stress in child development. *Current Directions in Psychological Science, 13*, 71–74.

DiPietro, J. A., Bornstein, M. H., Costigan, K. A., Pressman, E. K., Hahn, C. S., Painter, K., et al. (2002). What does fetal movement predict about behavior during the first two years of life? *Developmental Psychobiology, 40*, 358–371.

DiPietro, J. A., Bornstein, M. H., Hahn, C. S., Costigan, K. A., & Achy-Brou, A. (2007). Fetal heart rate and variability: Stability and prediction to developmental outcomes in early childhood. *Child Development, 78*, 1788–1798.

DiPietro, J. A., Caulfield, L. E., Costigan, K. A., Merialdi, M., Nguyen, R. H. N., Zavaleta, N., et al. (2004). Fetal neurobehavioral development: A tale of two cities. *Developmental Psychology, 40*, 445–456.

DiPietro, J. A., Christensen, A., & Costigan, K. A. (2008). The Pregnancy Experience Scale-Brief version. *Journal of Psychosomatic Obstetrics & Gynecology, 29*, 262–267.

DiPietro, J. A., Costigan, K. A., & Pressman, E. K. (1999). Fetal movement detection: Comparison of the Toitu actograph with ultrasound from 20 weeks gestation. *Journal of Maternal-Fetal & Neonatal Medicine, 8*, 237–242.

DiPietro, J. A., Ghera, M. M., Costigan, K. A., & Hawkins, M. (2004). Measuring the ups and downs of pregnancy. *Journal of Psychosomatic Obstetrics and Gynecology, 25*, 189–201.

DiPietro, J. A., Hilton, S. C., Hawkins, M., Costigan, K. A., & Pressman, E. K. (2002). Maternal stress and affect influence fetal neurobehavioral development. *Developmental Psychology, 38*, 659–668.

DiPietro, J. A., Hodgson, D. M., Costigan, K. A., Hilton, S. C., & Johnson, T. R. B. (1996). Development of fetal movement–fetal heart rate coupling from 20 weeks through term. *Early Human Development, 44*, 139–151.

DiPietro, J. A., Irizarry, R. A., Hawkins, M., Costigan, K. A., & Pressman, E. K. (2001). Cross-correlation of fetal cardiac and somatic activity as an indicator of antenatal neural development. *American Journal of Obstetrics and Gynecology, 185*, 1421–1428.

DiPietro, J. A., Kivlighan, K. T., Costigan, K. A., & Laudenslager, M. L. (2009). Fetal motor activity is associated with maternal cortisol. *Developmental Psychobiology, 51*, 505–512.

DiPietro, J. A., Novak, M. F., Costigan, K. A., Atella, L. D., & Reusing, S. P. (2006). Maternal psychological distress during pregnancy in relation to child development at age two. *Child Development, 77*, 573–587.

Dubowitz, L., Mercuri, E., & Dubowitz, V. (1998). An optimality score for the neurologic examination of the term newborn. *Journal of Pediatrics, 133*, 406–416.

Eaton, W., McKeen, N., & Campbell, D. (2001). The waxing and waning of movement: Implications for psychological development. *Developmental Review, 21*, 205–223.

Eaton, W. O., & Saudino, K. J. (1992). Prenatal activity level as a temperament dimension? Individual differences and developmental functions in fetal movement. *Infant Behavior and Development, 15*, 57–70.

Eggermont, J. J., & Salamy, A. (1988). Development of ABR parameters in a preterm and a term born population. *Ear and Hearing, 9*, 283–289.

Ellman, L. M., Schetter, C. D., Hobel, C. J., Chicz-DeMet, A., Glynn, L. M., & Sandman, C. A. (2008). Timing of fetal exposure to stress hormones: Effects on newborn physical and neuromuscular maturation. *Developmental Psychobiology, 50*, 232–241.

Field, T., Diego, M., Hernandez-Reif, M., Gil, K., & Vera, Y. (2005). Prenatal maternal cortisol, fetal activity, and growth. *International Journal of Neuroscience, 115*, 423–429.

Folsom, R. C., & Wynne, M. K. (1987). Auditory brainstem responses from human adults and infants: Wave V tuning curves. *Journal of the Acoustical Society of America, 81*, 412–417.

Freeman, R. K., Garite, T. J., & Nageotte, M. P. (1991). Physiologic basis of fetal monitoring. In R. K. Freeman, T. J. Garite, & M. P. Nageotte (Eds.), *Fetal heart rate monitoring* (2nd ed., pp. 7–20). Baltimore: Williams & Wilkins.

Fujioka, T., Fujioka, A., Tan, N., Chowdhury, G., Mouri, H., Sakata, Y., et al. (2001). Mild prenatal stress enhances learning performance in the non-adopted rat offspring. *Neuroscience, 103*, 301–307.

Gingras, J. L., & O'Donnell, K. J. (1998). State control in the substance-exposed fetus: I. The fetal neurobehavioral profile: An assessment of fetal state, arousal, and regulation competency. *Annals of the New York Academy of Sciences, 846*, 262–276.

Groome, L., Swiber, M., Holland, S., Bentz, L., Atterbury, J., & Trimm, R. (1999). Spontaneous motor activity in the perinatal infant before and after birth: Stability in individual differences. *Developmental Psychobiology, 35*, 15–24.

Gutteling, B., de Weerth, C., Willemsen-Swinkels, S., Huizink, A., Mulder, E., Visser, G., et al. (2005). The effects of prenatal stress on temperament and problem behavior of 27-month-old toddlers. *European Child and Adolescent Psychiatry, 14*, 41–51.

Hepper, P. G. (1995). Fetal behavior and neural functioning. In J. P. Lecanuet, W. P. Fifer, N. A. Krasnegor, & W. P. Smotherman (Eds.), *Fetal development: A psychobiological perspective* (pp. 405–417). Hillsdale, NJ: Erlbaum.

Hepper, P. G., & Shahidullah, S. (1992). Habituation in normal and Down's syndrome fetuses. *Quarterly Journal of Experimental Psychology, 44*, 305–317.

Horimoto, N., Koyangi, T., Maeda, H., Satoh, S., Takashima, T., Minami, T., et al. (1993). Can brain impairment be detected by in utero behavioural patterns? *Archives of Disease in Childhood, 69*, 3–8.

Huether, G. (1998). Stress and the adaptive self-organization of neuronal connectivity during early childhood. *International Journal of Neuroscience, 16*, 297–306.

Huizink, A., Robles de Medina, P., Mulder, E., Visser, G., & Buitelaar, J. (2002). Psychological measures of prenatal stress as predictors of infant temperament. *Journal of the American Academy of Child and Adolescent Psychiatry, 41*, 1078–1085.

Jiang, Z. D., Liu, X. Y., Shi, B. P., Lin, L., Bu, C. F., & Wilkinson, A. R. (2008). Brainstem auditory outcomes and correlation with neurodevelopment after perinatal axphyxia. *Pediatric Neurology, 39*, 189–195.

Jiang, Z. D., Xu, X., Brosi, D. M., Shao, X. M., & Wilkinson, A. (2007). Sub-optimal function of the auditory brainstem in term infants with transient low Apgar scores. *Clincal Neurophysiology, 118*, 1088–1096.

Johnson, T. R. B., Besinger, R. E., Thomas, R. L., Strobino, D. M., & Niebyl, J. R. (1992). Quantitative and qualitative relationships between fetal heart rate accelerations and fetal movement. *Journal of Maternal-Fetal & Neonatal Medicine, 1*, 251–253.

Kainer, F., Prechtl, H., Engele, H., & Einspieler, C. (1997). Assessment of the quality of general movements in fetuses and infants of women with type-I diabetes mellitus. *Early Human Development, 50*, 13–25.

Khedr, E., Farghaly, W., El-DinAmry, E., & Osman, A. (2004). Neural maturation of breastfed and formula-fed infants. *Acta Paediatrica, 93*, 734–738.

Krasnegor, N., Fifer, W., Maulik, D., McNellis, D., Romero, R., & Smotherman, W. (1998). Fetal behavioral development: A transdisciplinary perspective for assessing fetal well-being and predicting outcome. *Prenatal and Neonatal Medicine, 3*, 185–190.

Laplante, D., Barr, R., Brunet, A., DuFort, G., Meaney, M., Saucier, J., et al. (2004). Stress during pregnancy affects general intellectual and language functioning in human toddlers. *Pediatric Research, 56*, 400–410.

Litscher, G. (1995). Continuous auditory evoked potential measurements during nocturnal sleep. *International Journal of Neuroscience, 87*, 113–117.

Littleton, H., Breitkopf, C., & Berenson, A. (2007). Correlates of anxiety symptoms during pregnancy and association with perinatal outcomes: A meta-analysis. *American Journal of Obstetrics and Gynecology, 196*, 424–432.

Maeda, K., Morokuma, S., Yoshida, S., Ito, T., Pooh, R., & Serizawa, M. (2006). Fetal behavior analyzed by ultrasonic actocardiogram in cases with central nervous system lesions. *Journal of Perinatal Medicine, 34*, 398–403.

Maeda, K., Tatsumura, M., & Utsu, M. (1999). Analysis of fetal movements by doppler actocardiogram and fetal B-mode imaging. *Clinics in Perinatology, 26*, 829–851.

Martin, C. (1978). Regulation of the fetal heart rate and genesis of FHR patterns. *Seminars in Perinatology, 2*, 131–146.

Meek, L., Burda, K., & Paster, E. (2000). Effects of prenatal stress on development in mice: Maturation and learning. *Physiology and Behavior, 71*, 543–549.

Mohler, E., Parzer, P., Brunner, R., Wiebel, A., & Resch, F. (2006). Emotional stress in pregnancy predicts human infant reactivity. *Early Human Development, 82*, 731–737.

Moller, A. R., & Jannetta, P. J. (1982). Evoked potentials from the inferior colliculus in man. *Electroencephalography and Clinical Neurophysiology, 53*, 612–620.

Montandon, P. B., Cao, M. H., Engel, R. T., & Grajew, T. (1979). Auditory nerve and brainstem responses in the newborn and in preschool children. *Acta Oto-Laryngologica, 87*, 279–286.

Mueller, B. R., & Bale, T. L. (2008). Sex-specific programming of offspring emotionality after stress early in pregnancy. *Journal of Neuroscience, 28*, 9055–9065.

Mulder, E. J., Morssink, L. P., van der Schee, T., & Visser, G. H. (1998). Acute maternal alcohol consumption disrupts behavioral state organization in the near term fetus. *Pediatric Research, 44*, 774–779.

Nijhuis, I. J. M., & ten Hof, J. (1999). Development of fetal heart rate and behavior: Indirect measures to assess the fetal nervous system. *European Journal of Obstetrics, Gynecology, and Reproductive Biology, 87*, 1–2.

Nijhuis, I. J. M., ten Hof, J., Mulder, E. J., Nijhuis, J. G., Narayan, H., Taylor, D., et al. (2000). Fetal heart rate in relation to its variation in normal and growth retarded fetuses. *European Journal of Obstetrics, Gynecology, and Reproductive Biology, 89*, 27–33.

O'Brien, P., Wheeler, T., & Barker, D. (Eds.). (1999). *Fetal programming: Influences on development and disease in later life*. London: RCOG Press.

Poblano, A., Belmont, A., Sosa, J., Ibarra, J., Vargas, A., Limon, G., et al. (2003). Amikacin alters auditory brainstem conduction time in newborns. *Journal of Perinatology, 31,* 237–241.

Porges, S. W. (1983). Heart rate patterns in neonates: A potential diagnostic window to the brain. In T. Field & A. Sostek (Eds.), *Infants born at risk: Physiological, perceptual, and cognitive processes* (pp. 3–22). New York: Grune & Stratton.

Prechtl, H. F. R. (1984). Continuity and change in early neural development. In H. Prechtl (Ed.), *Continuity in neural functions from prenatal to postnatal life* (Vol. Clinics in Developmental Medicine No. 94, pp. 1–15). Philadelphia: Lippincott.

Robles de Medina, P., Visser, G., Huizink, A., Buitelaar, J., & Mulder, E. (2003). Fetal behaviour does not differ between boys and girls. *Early Human Development, 73,* 17–26.

Romanini, C., & Rizzo, G. (1995). Fetal behaviour in normal and compromised fetuses: An overview. *Early Human Development, 43,* 117–131.

Roncagliolo, M., Garrido, M., Walter, T., Peirano, P., & Lozoff, B. (1998). Evidence of altered central nervous system development in infants with iron deficiency anemia at 6 mo: Delayed maturation of auditory brainstem responses. *American Journal of Clinical Nutrition, 68,* 683–690.

Rothenberg, S., Poblano, A., & Schnaas, L. (2000). Brainstem auditory evoked response at five years and prenatal and postnatal blood lead. *Neurotoxicology and Teratology, 22,* 503–510.

Rotteveel, J. J., de Graaf, R., Colon, E. J., Stegeman, D. F., & Visco, Y. M. (1987). The maturation of the central auditory conduction in preterm infants until three months post term. II. The auditory brainstem responses (ABRs). *Hearing Research, 26,* 21–35.

Sandman, C. A., Wadhwa, P., Hetrick, W., Porto, M., & Peeke, H. (1997). Human fetal heart rate dishabituation between thirty and thirty-two weeks gestation. *Child Development, 68,* 1031–1040.

Shonkoff, J. P. (2006). A promising opportunity for developmental and behavioral pediatrics at the interface of neuroscience, psychology, and social policy: Remarks on receiving the 2005 C. Anderson Aldrich award. *Pediatrics, 118,* 2187–2191.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis.* New York: Oxford University Press.

Smotherman, W. P., & Robinson, S. R. (1987). Prenatal influences on development: Behavior is not a trivial aspect of fetal life. *Development and Behavioral Pediatrics, 8,* 171–175.

Sontag, L. W., & Richards, T. W. (1938). Studies in fetal behavior: I. Fetal heart rate as a behavioral indicator. *Monographs of the Society for Research in Child Development, 3*(4, Serial No. 17), 1–67.

Spielberger, C. (1983). *Manual for the State-Trait Anxiety Inventory (Form Y).* Palo Alto, CA: Mind Garden.

Tanner, J. M. (1978). *Foetus into man: Physical growth from conception to maturity.* Cambridge, MA: Harvard University Press.

Timor-Tritsch, I. E., Dierker, L. J., Zador, I., Hertz, R. H., & Rosen, M. G. (1978). Fetal movements associated with fetal heart rate accelerations and decelerations. *American Journal of Obstetrics and Gynecology, 131,* 276–280.

Unay, B., Sarici, S., Ulas, U., Akin, R., Alpay, F., & Gokcay, E. (2004). Nutritional effects on auditory brainstem maturation in healthy term infants. *Archives of Disease in Childhood, 89,* F177–F179.

Van den Bergh, B., Mennes, M., Oosterlaan, J., Stevens, V., Stiers, P., Marcoen, A., et al. (2005). High antenatal maternal anxiety is related to impulsivity during performance on cognitive tasks in 14- and 15-year-olds. *Neuroscience and Biobehavioral Reviews, 29,* 259–269.

Vintzileos, A. M., Campbell, W. A., & Nochinson, D. (1986). Relation between fetal heart rate accelerations, fetal movements, and fetal breathing movements. *American Journal of Perinatology, 3,* 16–18.

Weinstock, M. (2001). Alterations induced by gestational stress in brain morphology and behavior of the offspring. *Progress in Neurobiology, 65,* 427–451.

Yamasaki, M., Shono, H., Oga, M., Ito, Y., Shimomura, K., & Sugimori, H. (1991). Changes in auditory brainstem responses of normal neonates immediately after birth. *Biology of the Neonate, 60,* 92–101.

Young, J. (2002). Programming of sympathoadrenal function. *Trends in Endocrinology and Metabolism, 13,* 381–385.

Child Development, January/February 2010, Volume 81, Number 1, Pages 131–148

The Timing of Prenatal Exposure to Maternal Cortisol and Psychosocial Stress Is Associated With Human Infant Cognitive Development

Elysia P. Davis and Curt A. Sandman
University of California, Irvine

The consequences of prenatal maternal stress for development were examined in 125 full-term infants at 3, 6, and 12 months of age. Maternal cortisol and psychological state were evaluated 5 times during pregnancy. Exposure to elevated concentrations of cortisol early in gestation was associated with a slower rate of development over the 1st year and lower mental development scores at 12 months. Elevated levels of maternal cortisol late in gestation, however, were associated with accelerated cognitive development and higher scores at 12 months. Elevated levels of maternal pregnancy-specific anxiety early in pregnancy were independently associated with lower 12-month mental development scores. These data suggest that maternal cortisol and pregnancy-specific anxiety have programming influences on the developing fetus.

The prenatal period is a time of rapid change during which fetal organs and organ systems are forming and are vulnerable to both organizing and disorganizing influences. These influences on the fetus have been described as programming; the process by which a stimulus or insult during a vulnerable developmental period has a long-lasting or permanent effect. The effects of programming are dependent on the timing of the exposure and on the developmental stage of organ systems. There is convincing support for fetal programming of adult health outcomes; however, the evidence comes primarily from retrospective studies that rely on birth phenotype (e.g., small size at birth or preterm delivery) as an index of fetal development (Barker, 1998, 2002). It is unlikely, however, that birth phenotype alone is the cause of subsequent health outcomes. Birth phenotype, instead, reflects fetal adaptation to exposures that shape the structure and function of physiological systems that underlie health and disease risk (Gluckman & Hanson, 2004; Morley, Blair, Dwyer, & Owens, 2002). One emerging risk factor for health outcomes resulting from fetal programming is prenatal exposure to maternal stress signals. The purpose of the present study was to investigate the programming influence of biological and psychosocial indicators of prenatal maternal stress for fetal development and to evaluate the effects of timing of exposure to stress on infant development.

Fetal Programming: The Role of Glucocorticoids (GCs)

For a number of reasons, GCs have been proposed as a primary candidate for fetal programming. GCs cortisol in humans, are steroid hormones that play a critical role in normal development and are the end product of the hypothalamic–pituitary–adrenal (HPA) axis, one of the body's major stress responsive systems. HPA axis activity is regulated by the release of hypothalamic corticotrophin-releasing hormone (CRH), which stimulates the biosynthesis and release of adrenocorticotropic hormone (ACTH). Release of ACTH from the pituitary into the blood stream triggers cortisol production and release from the adrenal cortex. Cortisol is released into the general circulation and has effects on nearly every organ and tissue in the body (Munck, Guyre, & Holbrook, 1984). In human pregnancy, regulation of the HPA axis changes dramatically with the production and release of CRH from the placenta. In contrast to the role of cortisol in the negative feedback regulation of the HPA axis, cortisol stimulates placental CRH production resulting in a positive feedback loop that allows for the simultaneous increase of CRH,

---

This research was supported by a grant from the NIH (NS-41298). The authors wish to thank the families who participated in this project. The assistance of Cheryl Crippen and Carol Holliday is gratefully acknowledged.

Correspondence concerning this article should be addressed to Elysia Davis, 333 City Boulevard West, Suite 1200, Orange, CA 92868. Electronic mail may be sent to edavis@uci.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8101-0009

ACTH, and cortisol in the maternal and fetal compartments over the course of gestation (King, Nicholson, & Smith, 2001; Petraglia, Florio, Nappi, & Genazzani, 1996).

Maternal cortisol increases two- to fourfold over the course of normal gestation (Mastorakos & Ilias, 2003; Sandman et al., 2006). Fetal exposure to the increasing concentrations of maternal cortisol is regulated by a placental enzyme, 11β-hydroxysteroid dehydrogenase type 2 (11β-HSD2), which oxidizes cortisol to its inactive form cortisone (Beitens, Bayard, Ances, Kowarski, & Migeon, 1973; Brown et al., 1996). Activity of placental 11β-HSD2 also increases as pregnancy advances, providing partial protection for the fetus from maternal cortisol during critical stages of development. Toward the end of pregnancy, however, activity of the placental enzyme decreases, allowing a larger proportion of maternal cortisol to reach the fetus (Giannopoulos, Jackson, & Tulchinsky, 1982; Murphy, Smith, Giles, & Clifton, 2006). The normal increase in maternal cortisol during gestation and the decrease in placental 11β-HSD2 activity at the end of pregnancy ensure that the fetus is exposed to sufficient levels of cortisol during the third trimester, which are important for maturation of the fetal lungs and for preparation of the fetus for delivery (Austin & Leader, 2000; Hacking, Watkins, Fraser, Wolfe, & Nolan, 2001). Because placental 11β-HSD2 is only a partial barrier, active maternal cortisol passes through the placenta, and fetal cortisol levels are significantly correlated with maternal levels throughout gestation (Gitau, Cameron, Fisk, & Glover, 1998; Gitau, Fisk, Teixerira, Cameron, & Glover, 2001). GC receptors are present throughout the central nervous system (Diorio, Viau, & Meaney, 1993; Jacobson & Sapolsky, 1991; Sanchez, Young, Plotsky, & Insel, 2000) and GCs easily pass through the blood–brain barrier (Zarrow, Philpott, & Denenberg, 1970). GCs play a critical role in normal brain development, which provides further evidence for GCs as a mechanism for programming the fetus (Matthews, 2000; Trejo, Cuchillo, Machin, & Rua, 2000; Welberg & Seckl, 2001).

GCs shape the development of the neural systems involved in the regulation of emotion and cognitive function. Furthermore, moderate increases in GCs have salutary effects on both brain development and behavioral regulation (Kapoor, Dunn, Kostaki, Andrews, & Matthews, 2006). For example, rodents exposed in the early postnatal period to moderate increases in GCs display decreased physiological and behavioral stress responses, persisting improvements on cognitive tasks assessing learning and memory, and increased neural plasticity (Catalani et al., 2002; Scaccianoce, Catalani, Lombardo, Consoli, & Angelucci, 2001; Trejo et al., 2000). Exposure to excess GCs, however, particularly during vulnerable periods, appears to be neurotoxic (Uno et al., 1990; Uno et al., 1994). Animal models have demonstrated that prenatal exposure to administration of high levels of synthetic GCs has profound and lasting consequences for offspring including increased stress reactivity (Coe & Lubach, 2005; Hauser et al., 2007), poorer performance on cognitive tasks (Brabham et al., 2000), and impaired brain development, particularly in terms of reduced hippocampal volume (Bruschettini, van den Hove, Gazzolo, Steinbusch, & Blanco, 2006; Coe & Lubach, 2005; Uno et al., 1994). The few studies that have assessed the effects of prenatal treatment with a high dose of synthetic GCs in humans (given to women at risk of preterm delivery for maturation of fetal lungs) have documented consequences for emotional disturbances in early childhood (French, Hagan, Evans, Godfrey, & Newnham, 1999), dysregulated stress responses in infancy (Davis, Townsend, et al., 2004; Davis et al., 2006), impaired memory in school-aged children (MacArthur, Howie, Dezoete, & Elkins, 1982) and neurodevelopmental delays in toddlers (Spinillo et al., 2004; Wapner et al., 2007).

There are only a handful of studies that have evaluated the role that endogenous maternal cortisol plays in shaping human fetal development. Prenatal exposure to elevated maternal cortisol has been shown to predict increased fussiness, negative behavior and fearfulness in infancy (Davis et al., 2007; de Weerth, van Hees, & Buitelaar, 2003), and greater cortisol reactivity in childhood (Gutteling, de Weerth & Buitelaar, 2005). Only one group has evaluated the effects of endogenous maternal cortisol during pregnancy on infant cognitive and motor development. Huizink, Robles de Medina, Mulder, Visser, and Buitelaar (2003) have shown that elevated prenatal maternal cortisol during the third trimester was associated with delays in mental development at 3 months and motor development at 3 and 8 months. A limitation of these studies with humans is that it is difficult to separate shared genetic effects on emotion and stress regulation from direct effects of maternal cortisol. However, in conjunction with animal models where random assignment is possible and human studies involving synthetic GC administration, these studies suggest that prenatal exposure to endogenous maternal cortisol has a programming influence on the developing nervous system.

Fetal Programming: The Role of Maternal Psychosocial Stress

During pregnancy, maternal psychosocial stress threatens the fetal nervous system and shortens the length of gestation. In humans, the most well-documented effects of exposure to maternal stress is on birth outcomes including preterm delivery (Copper et al., 1996; Dole et al., 2003; Glynn, Dunkel Schetter, Hobel, & Sandman, 2008; Glynn, Wadhwa, Dunkel Schetter, & Sandman, 2001; Hobel & Culhane, 2003). Less is known about the consequences of prenatal maternal stress for human development independent of birth outcome. Numerous studies with rodent models have shown pervasive detrimental effects of exposure to prenatal maternal stress for the offspring (Dickerson, Lally, Gunnel, Birkle, & Salm, 2005; Kapoor, Petropoulos, & Matthews, 2008; Maccari & Morley-Fletcher, 2007; Van den Hove et al., 2005). Similarly, exposure of nonhuman pregnant primates to behavioral stressors during the prenatal period is associated with enhanced stress reactivity (Clarke, Wittwer, Abbott, & Schneider, 1994), compromised neuromotor development (Schneider & Coe, 1993; Schneider, Coe, & Lubach, 1992), irritable temperament (Schneider, 1992), poor attention regulation (Schneider, Roughton, Koehler, & Lubach, 1999), and impaired brain development, specifically reduced neurogenesis in the dentate gyrus (Coe et al., 2003). Stress early in gestation had more serious consequences for offspring than later gestational stress (Mueller & Bale, 2007; Schneider et al., 1999), although not all studies have found an effect of timing of exposure (Coe et al., 2003). The ability to generalize to humans from animal models is limited by the vast differences in reproductive physiology, even in closely related species, such as humans and nonhuman primates (Power et al., 2006) and by the significant differences in the type of prenatal stress employed with different species (e.g., physical restraint in rodents vs. report of stress in humans).

A small but growing literature with human populations has established that prenatal exposure to elevated levels of maternal stress is associated with behavioral and emotional disturbances during infancy and childhood, after controlling for postpartum maternal psychological distress, including postpartum depression (Bergman, Sarkar, O'Connor, Modi, & Glover, 2007; Davis et al., 2007; Davis, Snidman, et al., 2004; Gutteling, de Weerth, & Buitelaar, 2005; Huizink, De Medina, Mulder, Visser, & Buitelaar, 2002; O'Connor, Heron, & Glover, 2002). Effects of prenatal maternal stress on cognitive and motor development are less clear. There is evidence that maternal self-report of elevated stress and anxiety during the prenatal period is associated with delayed infant cognitive and neuromotor development (Brouwers, van Baar, & Pop, 2001; Huizink et al., 2003) and that these deficits may persist into adolescence (Mennes, Stiers, Lagae, & Van den Bergh, 2006). However, not all studies have demonstrated such an association (Brouwers et al., 2001; DiPietro, Novak, Costigan, Atela, & Ruesing, 2006; Van den Bergh, 1990) and is possible that generalized self-report measures of psychological distress do not adequately characterize stress that is unique during pregnancy. Evidence is emerging that measures of pregnancy-specific stress are better than measures of generalized psychological distress for predicting developmental outcomes including preterm delivery (DiPietro, Ghera, Costigan, & Hawkins, 2004; Roesch, Dunkel-Schetter, Woo, & Hobel, 2004), fetal behavior (DiPietro, Hilton, Hawkins, Costigan, & Pressman, 2002), infant cognitive and motor development (DiPietro et al., 2006; Huizink et al., 2003), infant emotional regulation (DiPietro et al., 2006), and child brain development (Buss, Davis, Muftuler, Head, & Sandman, in press).

Summary

During the prenatal period, maternal signals of stress have a programming influence on the developing fetus and it is possible that these signals have different consequences as gestation advances. There are, however, few prospective studies with humans that have addressed this question. Given the vast changes in both maternal physiology and fetal development during the prenatal period, characterization of the effects of prenatal maternal cortisol and psychosocial stress for development requires multiple longitudinal assessments during gestation. The goals of the current study were to determine, first, whether maternal cortisol and maternal and psychological stress during pregnancy were associated with cognitive and motor development in the offspring independent of postpartum influences, and second, whether there was a sensitive period for these effects on development.

Method

Participants

Study participants included mother–infant pairs from an ongoing longitudinal study of prenatal stress and development. Women with singleton pregnancies less than 16 weeks gestational age (GA) were recruited from obstetric clinics in Southern California and followed longitudinally from early pregnancy to 12 months postpartum. Women who were English speaking, nonsmokers, over the age of 18, did not take steroid medication, and for whom there was no evidence for drug or alcohol use during pregnancy were eligible for participation in this study. The current sample included the first 125 full-term infants (60 girls and 65 boys) and their mothers to complete the 12-month assessment. Women gave informed consent for all aspects of the protocol, which was approved by the Institutional Review Board for Protection of Human Subjects. Descriptive information for the study sample is shown in Table 1. All infants were full term at birth (*M* GA = 39.2 weeks, *SD* = 1.1 weeks; *M* weight = 3436 g, *SD* = 447 g). Infants in this sample were stable at the time of delivery and had a median 5-min Apgar score of 9 (range = 8–10).

Procedures

Maternal saliva samples were collected for cortisol analyses and maternal psychological state (state anxiety, perceived stress, and pregnancy-specific anxiety) was assessed at five intervals during pregnancy (Time 1: 15.1 ± 0.89; Time 2: 19.1 ± 0.93; Time 3: 25.4 ± .95; Time 4: 30.8 ± 0.69; Time 5: 36.6 ± 0.61 weeks of gestation). Maternal depression was assessed at the first four visits. Infant mental and psychomotor development was evaluated at 3 (*M* = 3.1, *SD* = 0.27), 6 (*M* = 6.2, *SD* = 0.3), and 12 (*M* = 11.9, *SD* = 0.3) months of age. At the three postnatal assessments, maternal state anxiety, perceived stress, depression, and parenting-specific stress were measured. The postpartum assessments of maternal psychological state were included to control for the potential influence of maternal affect during the postnatal period on infant development. Additionally, prenatal medical history and obstetric risk were obtained from extensive medical interviews conducted by a research nurse and through review of prenatal and hospital medical records.

Table 1
Demographic Information for the Study Sample

	Study sample (*N* = 125)
Maternal age at delivery	29.9[a]
Married (%)	79
Primiparous (%)	48
Education (%)	
High school or equivalent	95
College graduate	46
Annual household income (%)	
$0–$30,000	17
$30,001 to $60,000	26
$60,001 to $100,000	33
Over $100,000	24
Race/ethnicity (%)	
Non-Hispanic White	50
Hispanic	30
Asian	10

[a] *SD* = 5.3, range = 19–41.

Measures

Salivary cortisol assessment. Saliva samples were collected in the early afternoon, at least 1 hr after the participant had eaten (mean time of day ranged from 13:19 to 13:30 across the five assessment intervals) using a Salivette sampling device (Sarstedt, Numbrecht, Germany). Within this restricted range of sample collection time, time of day was not associated with cortisol levels (all *r*s < .15, all *p*s > .10).

Saliva samples were spun and stored at −70°C until assayed. Thawed samples were centrifuged at 1500 × g for 15 min before assay. Salivary cortisol levels were determined by a competitive luminescence immunoassay (LIA; IBL-America, Minneapolis, MN) with reported detection limits of 0.015 μg/dl. The cross reactivity of the assay was < 2.5% with cortisone, prednisone, and corticosterone and < 0.1% with other naturally occurring steroids. The intra- and interassay coefficients of variance are 5.5% and 7.6%, respectively. Data reduction for the LIA assay was done by an automated four-parameter logistics computer program (software Mikro Win 2000; Berthold Microplate Luminometer, Berthold Detection Systems GmbH; Pforzheim, Germany). One maternal sample from the second prenatal visit was removed because it was more than 4 *SD* above the mean. All samples were assayed in duplicate and averaged. A square root transformation was implemented to reduce skewedness. All values reported in the text, tables and graphs are raw μg/dl levels to facilitate interpretation.

Maternal psychological assessments. Generalized or nonspecific stress was evaluated using the 12-item version of Cohen's Perceived Stress Scale (PSS; Cohen, Kamarck, & Mermelstein, 1983). The PSS evaluated participants' feelings about how they were able to handle day-to-day problems and hassles, how often they felt nervous and stressed, and how often they felt things were going well. Responses were made on a 5-point Likert scale ranging from 1 (*never*) to 5 (*almost always*) and the final score could range from 12 to 60. This measure has been used extensively with nonpregnant and pregnant samples (Culhane et al., 2001; Glynn et al., 2008) and been shown to have both good internal consistency ($\alpha = .80$) and validity (Cohen et al., 1983).

The short form of the Center for Epidemiological Studies Depression Inventory was used to evaluate maternal depression (CES–D; Santor & Coyne, 1997). Responses to each of the nine items in this measure were recorded on a 4-point Likert scale with a range of 0–3. Anchor points, in terms of days per week, were *rarely or none of the time (less than 1 day)* to *most or all of the time (5–7 days)*. The final score could range from 0 to 27, with a higher score indicating greater impairment. This measure has been extensively used and published studies demonstrate both internal consistency ($\alpha = .84$) and validity of this measure (Santor & Coyne, 1997). The CES–D is a frequently used instrument for the study of depression in the general population and has been validated in pregnant samples (Marcus, Flynn, Blow, & Barry, 2003). Scores on the CES–D during pregnancy are associated with birth outcome and infant development (Davis et al., 2007; Lundy et al., 1999).

The 10-item State Anxiety subscale of the State-Trait Personality Inventory (STAI; Speilberger, 1983) was used to measure state anxiety. The STAI assessed the degree to which participants had experienced anxiety-related symptoms or emotions. Responses were made using a 4-point Likert scale ranging from 1 (*not at all*) to 4 (*very much*) and scores could range from a minimum of 10 to a maximum of 40. The STAI has been used for research purposes with both pregnant (Glynn et al., 2008; Rini, Dunkel-Schetter, Wadhwa, & Sandman, 1999) and nonpregnant samples. The STAI has good internal consistency with a Cronbach's alpha coefficient of .92 (Spielberger, 1983).

The 10-item Pregnancy Specific Anxiety (PSA) scale was used to measure a woman's feelings about her health during pregnancy, the health of her baby, and her feelings about labor and delivery. Answers were given on a 4-point Likert scale ranging from 1 (*not at all*) to 4 (*very much*) and included items such as: "I am fearful regarding the health of my baby," "I am concerned or worried about losing my baby," and "I am concerned or worried about developing medical problems during my pregnancy." The final score on this measure could range from 10 to 40. This reliable measure was specifically developed for use in pregnancy research and has good internal consistency (αs = .75–.85; Glynn et al., 2008; Rini et al., 1999).

The 36-item Parenting Stress Index (PSI) is a measure designed to identify stress within the parent–child relationship. Responses were given on a 5-point Likert scale. This measure includes items such as: "I find myself giving up more of my life to meet my children's needs than I ever expected" and "My child rarely does things for me that make me feel good." The total score could range from 36 to 180. The PSI has good internal consistency ($\alpha = .95$; Albidin, 1995).

Medical Risk and Obstetric History

An extensive structured medical interview was conducted by a research nurse at the participants' first prenatal visit to assess pregnancy history and to identify risk factors that might affect pregnancy outcome. At each subsequent visit, a follow-up interview was conducted to characterize pregnancy-related changes in maternal health. Maternal and infant medical records were reviewed to assess prenatal medical history and birth outcome. A binary score assessing the presence or absence of prenatal medical risk was derived (Hobel, 1982). The risk factors considered were infection, hypertension, diabetes, oligohydramnios, polyhydramnios, preterm labor, vaginal bleeding, placenta previa and anemia, in the index pregnancy. Additionally, birth order, GA at birth, birth weight, and Apgar scores were recorded.

Infant Development

Infant development was assessed using the Bayley Scales of Infant Development, 2nd ed. (BSID–II). Examiners were trained by a clinician with over 15 years of experience with the BSID and were directly supervised by a clinical psychologist. Videotaped assessments were reviewed monthly. Interrater reliability, calculated on 20% of the assessments at each age, was 95% at 3 and 6 months and 93% at 12 months. The BSID is a standardized developmental assessment (Bayley, 1993). Measures of development were obtained at each age using the Mental Developmental Index

(MDI) and Psychomotor Development Index (PDI). Sample items from the MDI and PDI at each age are presented in Table 2. Reliability coefficients on the BSID were .83, .92, and .88 for MDI and .84, .84, and .84 for PDI at 3, 6, and 12 months of age, respectively. Conventional scoring of the BSID created composite MDI and PDI scaled scores by summing the total number of items achieved, corrected for basal effects. The raw scores were converted to scaled scores by reference to a developmental table. Mean score raw scores and scaled scores for the MDI and PDI are presented in Table 3.

Data Analysis

Preliminary analyses were performed using regressions and *t* tests to identify sociodemographic (i.e., race/ethnicity, marital status, education, and household income), pregnancy-related (i.e., prenatal medical risk), and infant (i.e., GA at birth, sex, birth order, and ever breastfed) variables that might influence infant cognitive or motor development. The factors associated with infant outcomes with a p value of .10 or less were GA at birth (MDI and PDI; $p < .01$), maternal race ($p < .10$), prenatal medical risk (MDI only; $p = .10$), and birth order (PDI only; $p < .01$). None of the assessments of postnatal maternal psychological state (state anxiety, depression, or perceived stress) significantly predicted infant development (all $ps > .2$). Furthermore, maternal report of stress within the parent–child relationship was not significantly associated with infant development (all $ps > .2$). All variables that were associated with MDI or PDI in preliminary analyses (infant GA at birth, maternal race, birth order, and prenatal medical risk) were modeled as covariates. Although postnatal measures of psychological distress were not significantly associated with infant MDI or PDI at 3, 6, or 12 months, models were tested with and without these measures. Adjusting for postnatal maternal psychological distress did not change the statistical significance of the results in any of the analyses described below.

Hierarchical linear modeling (HLM) growth curve analyses (Raudenbush & Bryk, 2002; Singer & Willett, 2003) were used to describe both the trajectory of the prenatal indicators of maternal stress (maternal cortisol and psychosocial measures) across pregnancy and the trajectory of infant development as assessed with the BSID. HLM models were used to determine whether the patterns of prenatal endocrine and psychosocial stress measures were associated with individual differences in infant development. HLM, when used with repeated measures, treats the data in a hierarchical fashion with observations nested within persons. This approach allows variance to be modeled at multiple levels and provides several advantages over ordinary least squares regression (Raudenbush & Bryk, 2002). First, standard regression or analysis of variance (ANOVA) models are limited to one component of variability, the deviation of the individual from the group mean. In contrast, HLM includes the within-person variability assessed over time. Second, estimates of goodness of fit in modeling each individual's data are derived, and the most reliable data are given greater statistical weight. Third, HLM produces robust estimates despite missing values for the repeated dependent measure. Ninety-nine participants had complete data at all prenatal assessments. Twenty-three participants were missing data

Table 2
Sample Items From the Mental and Psychomotor Scales

	Mental Scale	Psychomotor Scale
3 months	Discriminates novel visual pattern	Elevates self by arms
	Eyes follow rod	Sits with support
	Reaches for suspended ring	Balances head
6 months	Lifts inverted cup	Uses whole hand to grasp pellet
	Looks for fallen spoon	Sits alone
	Pulls string to secure ring	Turns from back to stomach
12 months	Puts one cube in cup	Walks alone
	Responds to spoken request	Throws ball
	Places one puzzle piece	Squats briefly

Table 3
Descriptive Information (Mean and Standard Deviation) for Infant BSID Scores

	Infant age		
	3	6	12
Mental development raw score	30.4 (3.3)	61.2 (4.3)	84.9 (4.3)
MDI scaled score	92.8 (5.9)	96.6 (8.8)	93.1 (10.3)
Psychomotor development raw score	23.1 (2.8)	40.0 (4.2)	64.1 (4.3)
PDI scaled score	93.7 (8.6)	97.3 (12.9)	97.1 (16.8)

Note. BSID = Bayley Scales of Infant Development, 2nd ed. (BSID–II); MDI = Mental Developmental Index; PDI = Psychomotor Development Index.

from one visit and 3 participants were missing data from two visits. One hundred twelve participants had complete data at all postnatal assessments. Eight participants were missing data from one visit and 5 participants were missing data from two visits. In HLM models cases with complete data are weighted more heavily, but all cases are included in the estimation of effects. Finally, HLM uses precise measures of timing (i.e., gestational week) of data collection rather than nominal estimates of assessment intervals.

Using HLM, two sets of analyses were performed. The first analysis was implemented to examine whether prenatal maternal stress measures at a given prenatal assessment predicted the trajectory of infant development over the first postnatal year. In these analyses, both mental and motor performance at a given age (3, 6, and 12 months) and developmental trajectories (slope) over the first postnatal year were assessed. To characterize developmental trajectories, raw data (rather than scaled scores) from the mental and psychomotor scales were used in these analyses. Data were modeled at the exact age in weeks that infants were tested to account for variability in assessment age. The second set of analyses assessed the profile or trajectory of prenatal maternal stress measures that was associated with infant performance at 12 months using scaled scores determined based on developmental norms for the BSID.

Predicting Trajectories of Infant Development

To determine whether prenatal maternal stress measures predicted the trajectory of infant development across the first postnatal year, two-level models were constructed. For these analyses, raw data from the mental and psychomotor scales were used so that the trajectory of development could be assessed. Level 1 variables, or those evaluated longitudinally across the first postnatal year, included: infant mental and psychomotor development (raw data) and infant age in weeks entered as continuous variables. In all cases, a linear model best fit the data. Maternal stress measures (cortisol, perceived stress, state anxiety, pregnancy-specific anxiety, and depression) measured at each of the five prenatal assessments and entered as continuous variables were modeled at Level 2. Prenatal predictors were assessed at each of the five prenatal assessment time points. Time-invariant variables that satisfied criterion for inclusion as covariates (GA at birth, race/ethnicity [Hispanic: 0 = no, 1 = yes; White: 0 = no, 1 = yes], prenatal medical risk [MDI only; 0 = no, 1 = yes], and birth order [PDI only; 0 = firstborn, 1 = later born]) were included at Level 2. For each Level 1 variable, the models were tested for differences in slope and intercept at 3, 6, and 12 months of age.

Patterns of Maternal Biosocial Stress Associated With Infant Outcomes

At the infant ages where significant associations were found between predictor and outcome variables, a second set of analyses was performed to characterize the profile of maternal stress across pregnancy that was associated with infant outcomes at a given age. In all cases, predictor and outcome variables were modeled as continuous variables. For these analyses, two-level models were constructed to evaluate maternal cortisol and psychosocial stress trajectories across pregnancy. Level 1 variables, or those evaluated longitudinally across the five prenatal assessment days, included: maternal cortisol, psychosocial measures (perceived stress, state anxiety, pregnancy-specific anxiety, and depression), and gestational week. A linear model was applied in all cases except for pregnancy-specific anxiety, where a quadratic function was a better fit, $\chi^2(3) = 12.1, p < .01$.

BSID scores (MDI and PDI), entered as continuous variables, were the outcomes of interest and were modeled at Level 2. Time-invariant variables that satisfied criterion for inclusion as covariates (GA at birth, race/ethnicity [Hispanic: 0 = no, 1 = yes; White: 0 = no, 1 = yes], prenatal medical risk [MDI only; 0 = no, 1 = yes], and birth order [PDI only; 0 = firstborn, 1 = later born]) as well as postnatal measures of psychosocial distress were included at Level 2. For each Level 1 variable, the model was tested for differences in slope and intercept at all time points within the range of actual endocrine and psychosocial assessments (13–38 weeks of gestation). Individual slope and intercept values were retained from significant models. The values that were most strongly associated with infant outcomes were entered into stepwise and hierarchical regression models to determine whether maternal psychosocial and endocrine stress measures jointly or independently predicted BSID scores.

Results

Infant Development

The unconditional growth model was significant indicating that mental ($t = 105.2$, $p < .001$) and psychomotor ($t = 108.3$, $p < .001$) performance (raw scores) improved across the first postnatal year (see Table 3). As expected, scaled scores calculated with reference to developmental norms, did not increase ($p > .30$). Table 4 shows the correlation between mental and psychomotor development scaled scores across the three assessment time points. As illustrated, modest stability is seen in BSID scores with a stronger association among temporally adjacent measures.

Maternal Cortisol

The unconditional growth model was significant, indicating that, as expected, salivary cortisol levels increased across gestation ($t = 13.6$, $p < .001$). Salivary cortisol levels increased from .27 at 15 gestational weeks to .61 at 37 gestational weeks. Table 5 shows the correlation between cortisol measures across pregnancy. Moderate correlations were seen between temporally adjacent assessments.

Maternal Psychological State

Maternal perceived stress ($Ms = 20.1$ to 21.6), depression ($Ms = 13.7$ to 15.1), and state anxiety ($Ms = 17.7$ to 18.6) did not significantly change over the course of pregnancy ($ps = .06$ to .30). Pregnancy-specific anxiety changed significantly across gestation ($t = 3.4$, $p < .01$). Levels were highest early in pregnancy and decreased through mid-gestation. Mean PSA scores ranged from 17.3 to 18.7.

Each of the four psychosocial measures was stable over pregnancy. Within each measure intercorrelations ranged from .51 to .75 during pregnancy. Not only were scores on the psychosocial measures stable during the prenatal period, but as expected, concurrent measures of state anxiety, perceived stress, and depression were intercorrelated (average $rs = .66$, $ps < .001$). The associations between these measures and PSA were lower but also significant (average $rs = .41$, $ps < .001$).

During the postnatal period, levels of state anxiety ($Ms = 16.8$ to 17.5), perceived stress ($Ms = 20.1$ to 21.1), and depression ($Ms = 12.8$ to 12.9) were relatively stable ($rs = .45$ to .68) and concurrent measures were intercorrelated (average $rs = .67$). Mean levels of parenting stress ranged from 64.2 to 66.6 and parenting stress was relatively stable over the first postnatal year ($rs = .48$ to .67).

Prenatal Maternal Cortisol and Maternal Psychological State

Data indicate that maternal cortisol was not significantly associated with maternal anxiety, depression, or pregnancy-specific anxiety at any of the five prenatal assessment time points (all $rs < .18$). Prenatal maternal perceived stress was not associated with maternal cortisol at any of the prenatal assessment time points (all $rs < .13$) with the exception of the 36+ weeks' GA assessment. At this time point, maternal cortisol was negatively associated with perceived stress levels, $r(119) = -.21$, $p < .05$.

Maternal Prenatal Stress Measures and Trajectories of Infant Development

We first examined whether the rate of infant development over the first postnatal year was associated with prenatal cortisol at each assessment interval: 15, 19, 25, 31, and 37 weeks. For these analyses, raw scores (rather than scaled scores) were used so that developmental changes could be evaluated. Data indicate that the level of maternal cortisol both early and late in gestation predicted infant mental development. Lower maternal cortisol at the first prenatal assessment (15 weeks) predicted accelerated infant mental

Table 4
Intercorrelations Among MDI and PDI Scaled Scores Across the First Postnatal Year

	3-month MDI	3-month PDI	6-month MDI	6-month PDI	12-month MDI	12-month PDI
3-month MDI	—	.41**	.23*	.08	.06	.13
3-month PDI	—	—	.16	.22*	.15	.26**
6-month MDI	—	—	—	.61**	.27**	.40**
6-month PDI	—	—	—	—	.37**	.55**
12-month MDI	—	—	—	—	—	.43**

Note. MDI = Mental Developmental Index; PDI = Psychomotor Development Index.
$^{*}p < .05$. $^{**}p < .01$.

Table 5

Intercorrelations Among Prenatal Maternal Cortisol Samples

	Gestational week			
Gestational week	19	25	31	37
15	.21*	.34**	.03	.05
19	—	.22*	.11	.09
25	—	—	.07	.02
31	—	—	—	.39**

$^*p < .05$. $^{**}p < .01$.

development (steeper slope of raw scores) across the first postnatal year ($t = -2.5$, $p < .01$), resulting in enhanced cognitive functioning at 12 months of age ($t = 2.4$, $p < .05$). Furthermore, higher maternal cortisol late in gestation (37 weeks) additionally predicted accelerated infant development (a steeper slope of raw MDI scores) over the first postnatal year ($t = 1.9$, $p < .05$), resulting in a tendency for higher mental development raw scores at 12 months ($t = 1.6$, $p = .10$). Maternal prenatal cortisol at 19, 25, and 31 week assessments did not predict the trajectory of infant development assessed using mental development raw scores. No measure of maternal cortisol at any time period predicted psychomotor development raw scores. None of the measures of maternal psychological distress (perceived stress, state anxiety, pregnancy-specific anxiety, and depression) were significantly associated with infant development using raw mental or psychomotor scores. These data suggest that the pattern of maternal cortisol during gestation predicts the trajectory of infant mental development over the first postnatal year resulting in significant differences in performance only at 12 months.

Patterns of Maternal Cortisol Associated With Infant Development

A second set of analyses were performed to identify the pattern of maternal cortisol that predicted optimal performance at 12 months. For these analyses, the scaled scores were used to allow for comparison to developmental norms. These models offer the additional advantage of assessing the slope or trajectory of maternal cortisol across gestation. As in the previous analyses, cortisol and BSID scores were modeled as continuous variables. As shown in Figure 1, infants scoring higher on the 12-month MDI were exposed to (a) lower maternal cortisol before 18 weeks GA (ts = −2.1 to −2.7, ps < .05), (b) an accelerated increase in maternal cortisol across gestation ($t = 3.7$, $p < .001$), and (c) higher maternal cortisol after 30 weeks GA (ts = 2.0 to 2.9, ps < .05). Interestingly, this profile of maternal cortisol additionally predicted whether or not 12-month-old infants received a score that would be categorized as delayed using standardized cutoff on the BSID (ps < .05). Because the intercepts that were the strongest predictors of a low MDI score were high maternal cortisol at 13 weeks GA and low maternal cortisol at 38 weeks GA, a stepwise regression model was implemented to determine whether early and late cortisol jointly or independently predicted MDI. The stepwise model indicated that, after entering covariates into the model, early and late cortisol independently predicted infant MDI (13-week GA intercept: $\Delta R^2 = .05$, $\beta = -.23$, $t = -2.5$, $p < .01$; 38-week GA intercept: $\Delta R^2 = .03$, $\beta = .17$, $t = 2.0$, $p < .05$). Together, early and late maternal cortisol accounted for 8% of the variance in infant cognitive performance, as assessed with the MDI. At 13 weeks of gestation, a 0.1 μg/dl increase in cortisol was associated with a 4-point decrease in MDI. At 38 weeks of gestation, a 0.1 μg/dl increase in cortisol was associated with a 2-point increase in MDI. PDI was not significantly associated with maternal cortisol levels (ps > .3).

To determine whether there was a profile of prenatal maternal cortisol that predicted infant

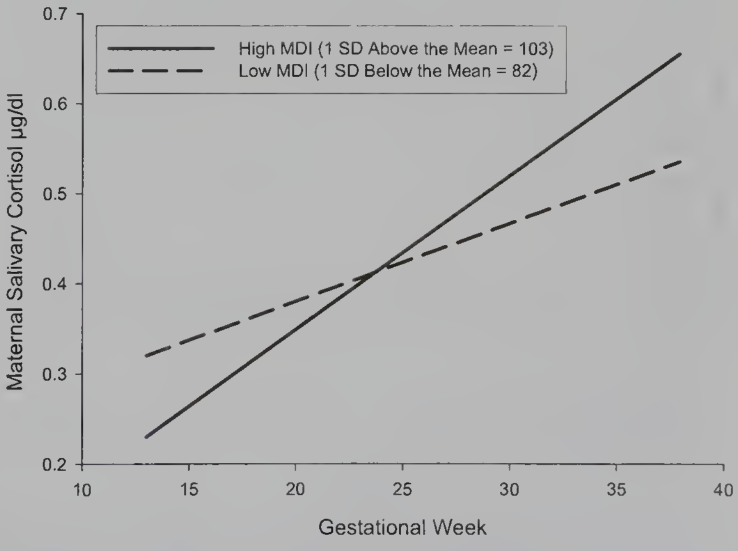


Figure 1. Infant performance on the Mental Developmental Index (MDI) at 12 months of age is associated with the trajectory of maternal cortisol across gestation. The top and bottom standard deviations for MDI are modeled with hierarchical linear modeling.

Note. Both maternal cortisol and MDI scores were analyzed as continuous variables. For graphing purposes, we have displayed the cortisol trajectory for mothers of infants with MDI scores 1 *SD* above or below the mean.

cognitive development, the slope of maternal cortisol across gestation was calculated for each woman. When the slope of maternal cortisol across pregnancy was added to the model it was demonstrated that cortisol slope was a stronger predictor of infant MDI ($\Delta R^2 = .06$, $\beta = .24$, $t = 2.8$, $p < .01$) than either early or late cortisol levels. Furthermore, after cortisol slope was entered into the model the effects of maternal cortisol early and late in pregnancy (intercepts at 13 or 38 weeks GA) were no longer significant ($ps > .3$). These data suggest that it is the profile of maternal cortisol across gestation that best predicts infant development.

Maternal Psychosocial Stress and Infant Development

A second set of HLM models were computed to examine associations between psychosocial indicators of maternal stress and infant development. Among all the measures of maternal distress (perceived stress, state anxiety, pregnancy-specific anxiety, and depression) only pregnancy-specific anxiety was significantly associated with infant development. Infant performance on the MDI was associated with both pregnancy-specific anxiety early in gestation and overall rate of change (shape of the curve) of pregnancy-specific anxiety across gestation ($t = 2.4$, $p < .05$). As shown in Figure 2, infants scoring low on the MDI had mothers with the profile of (a) high pregnancy-specific anxiety before 16 weeks GA ($ts = -2.6$ to -1.9, $ps < .05$) and (b) a steeper decline in pregnancy-specific anxiety through mid gestation (for rate of acceleration or instantaneous slope, $ts = 2.1$ to 2.9, $ps < .05$), such that no differences in level of pregnancy-specific anxiety were detected beyond 16 weeks of gestation. Maternal pregnancy-specific anxiety did not significantly predict whether infants were categorized as delayed using BSID criteria ($ps > .17$). Thirteen weeks GA was the time point at which level of pregnancy-specific anxiety most strongly predicted infant cognitive development. At 13 gestational weeks, a 5-point increase in PSA was associated with a 2-point decrease on the MDI. A stepwise regression was implemented to determine whether levels of pregnancy-specific anxiety at 13 weeks GA or the rate of change of pregnancy-specific anxiety over gestation best predicted infant scores on the MDI. After adjusting for covariates, overall rate of change of PSA was a stronger predictor of MDI than the 13-week GA intercept ($\Delta R^2 = .05$, $\beta = -.21$, $t = -2.4$, $p < .05$) and accounted for the variance associated with the 13-week GA intercept ($p = .24$).

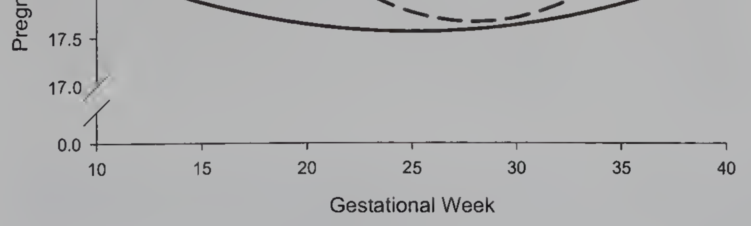


Figure 2. Elevated pregnancy specific anxiety early in gestation is associated with poorer performance on the Mental Developmental Index (MDI) at 12 months of age. The top and bottom standard deviations for MDI are modeled with hierarchical linear modeling.

Note. Both maternal pregnancy specific anxiety and MDI scores were analyzed as continuous variables. For graphing purposes, we have displayed the trajectory of pregnancy specific anxiety for mothers of infants with MDI scores 1 *SD* above or below the mean.

Influence of Maternal Psychological and Endocrine Factors on Infant Development

A hierarchical regression model was implemented to determine whether psychosocial (pregnancy-specific anxiety) and endocrine (cortisol) measures independently or jointly predicted portions of the variance in infant cognitive development. Furthermore, after adjusting for covariates, maternal cortisol measures (slope and intercepts, $\Delta R^2 = .07$, $p < .01$) and maternal PSA indicators (slope and intercept, $\Delta R^2 = .05$, $p < .05$) independently predicted MDI. Together these two measures accounted for 12% of the variance in infant performance on the MDI at 12 months of age.

Discussion

The current study provides further evidence that maternal cortisol and signals of maternal psychosocial stress play a role in the programming of the human fetus. Both maternal cortisol and maternal psychosocial stress (pregnancy-specific anxiety) were associated with the trajectory of infant development and influenced cognitive functioning at 1 year of age among healthy, full-term infants. Furthermore, the consequences of prenatal maternal stress were determined by the gestational period during which the fetus was exposed.

Prenatal Maternal Cortisol Influences Development

The present study suggests that prenatal maternal cortisol exerts a programming effect on human infant development. Maternal cortisol influenced the trajectory of infant development over the first postnatal year, resulting in differences in performance on the MDI at 12 months. Performance on the PDI was not associated with maternal cortisol.

As expected, maternal cortisol significantly increased across gestation. The pattern of prenatal maternal cortisol was associated with infant cognitive functioning. Prenatal maternal cortisol levels predicted the rate of infant cognitive development over the first postnatal year, resulting in significant differences in performance on the MDI at 12 months of age. Specifically, infants who scored high on the MDI at 12 months of age were exposed to (a) low concentrations of maternal cortisol in early pregnancy, (b) an accelerated trajectory of cortisol across gestation, and (c) high concentrations of cortisol toward the end of pregnancy. Thus, prenatal maternal cortisol had opposite effects on infant cognitive development depending on the timing of exposure. These findings are remarkably consistent with the role cortisol plays in the maturation of the human fetus. Early in pregnancy the fetus is protected from the naturally occurring increases in maternal cortisol by 11β-HSD2, an enzyme that converts cortisol into its inactive form. However, because 11β-HSD2 is only a partial barrier, excessive increases in maternal cortisol early in gestation will result in overexposure of the fetus with potentially detrimental consequences. In contrast, as pregnancy advances toward parturition exposure to elevated cortisol is necessary and beneficial for maturation of fetal organ systems, including the fetal lungs (Hacking et al., 2001). Fetal exposure to increased cortisol during the 3rd trimester is facilitated by the sharp decline in 11β-HSD2 activity which allows a greater proportion of maternal cortisol to cross the placental barrier (Giannopoulos et al., 1982; Murphy et al., 2006). The current data provide new evidence that the programming influence of maternal cortisol on the developing fetal nervous system is dependent on the timing of exposure. Exposure to elevated concentrations of cortisol early in gestation had deleterious effects on subsequent infant cognitive development, whereas exposure to elevated levels of maternal cortisol late in gestation was associated with enhanced performance on tasks assessing mental development. It is possible that effects were observed for the MDI and not PDI because areas of the brain that are critical for cognitive functioning such as the hippocampus and prefrontal cortex are more vulnerable to the effects of GCs (Lowy, 1994; Uno et al., 1994; Welberg, Seckl, & Holmes, 2001).

GCs, in addition to their critical role in lung maturation, play an important role in normal brain development and have been implicated in neuronal maturation and survival (Drake, Tang, & Nyirenda, 2007; Scaccianoce et al., 2001; Trejo et al., 2000). The association between increased concentrations of cortisol during the third trimester and improved cognitive functioning at 12 months of age is highly plausible given the role that GCs have been shown to play in regulating brain development in animal models. Although excessive prenatal administration of GCs causes both neural degeneration and impaired performance on cognitive tasks (Coe & Lubach, 2005; Uno et al., 1990; Uno et al., 1994), moderate increases in GCs facilitate brain development resulting in enhanced cognitive functioning (Casolini et al., 1997; Catalani et al., 2000; Scaccianoce et al., 2001). The current findings emphasize that within physiological limits during the prenatal period, the level of GCs associated with optimal development will be determined by the gestational period. Optimal development was observed in infants of mothers with low levels of maternal cortisol early and high levels late in pregnancy.

One other group that examined the effect of prenatal exposure to endogenous maternal cortisol for human infant cognitive development reported that elevated maternal cortisol during the third trimester was associated with poor performance on the MDI at 3 months of age (Huizink et al., 2003). The reason for the discrepancy between this report and the current study, where third trimester maternal cortisol was unrelated to 3-month BSID performance and positively related to 12-month BSID performance is unclear. There are, however, differences in the design of the two studies. The current investigation included longitudinal assessments of infants across the first postnatal year allowing us to evaluate the trajectory of infant cognitive development. Furthermore, at 3 months of age, performance on the BSID is highly influenced by infant behavioral state and later MDI scores are more stable than earlier scores (Bayley, 1993). The current study included a large sample of 125 mother–infant pairs and serial assessments of prenatal maternal cortisol levels at five distinct gestational periods. In this large sample with precise measures of maternal cortisol over the course of pregnancy, our results indicated that poor performance on the MDI at 12 months was associated with elevations in maternal cortisol early in pregnancy and not during the third trimester.

Prenatal Maternal Psychosocial Stress Influences Infant Development

Among all of the measures of prenatal maternal psychological distress (perceived stress, state anxiety, pregnancy-specific anxiety, and depression), only pregnancy-specific anxiety significantly predicted infant development. Maternal report of pregnancy-specific anxiety was highest early in pregnancy and declined as gestation advanced. Furthermore, both the level of pregnancy-specific anxiety early in gestation and the overall change in pregnancy-specific anxiety significantly predicted infant performance on the MDI, but not PDI, at 12 months of age. Pregnancy specific anxiety was not associated with infant development at earlier ages. A low infant MDI score at 12 months of age was associated with the maternal profile of high pregnancy-specific anxiety before 16 weeks GA and a steeper decrease in pregnancy-specific anxiety through mid gestation, at which point pregnancy-specific anxiety was not associated with infant development. It may be surprising that a steeper decline in pregnancy-specific anxiety was associated with poorer infant scores on the MDI. It is likely, however, that the steeper decline in pregnancy-specific anxiety reflects higher initial values; levels of pregnancy-specific anxiety did not differ later in gestation. It is not surprising that pregnancy relates stress changes across gestation. The stimuli that lead to pregnancy-specific stress appraisals are different early in gestation (e.g., fear of miscarriage) as compared to late in gestation (e.g., concern about labor and delivery). Pregnancy anxiety was not measured prior to 13 weeks GA and thus it remains possible that pregnancy anxiety earlier in pregnancy would have a greater effect. Further research is needed to better understand the way that pregnancy-specific stress changes across gestation as well as how pregnancy-specific stress is different from other types of stress. Importantly, the mother's pregnancy-specific anxiety predicted infant development even after controlling for actual medical risk and postnatal maternal stress. These data suggest that (a) signals of maternal psychosocial stress during the prenatal period may be transmitted to the fetus and have programming consequences for development, (b) the fetus is more vulnerable to the effects of these signals early in gestation, and (c) maternal stress related to pregnancy may have more profound implications for the fetus as compared to generalized maternal distress.

The finding that pregnancy-specific anxiety and not generalized distress measures predicts infant outcomes suggests that stress related to events, beliefs, or fears proximal to pregnancy are both novel and more relevant to the pregnant woman and therefore more consequential to the fetus and that this measure detects unique emotional distress. In contrast to the three generalized distress measures that evaluate stress related to generalized worry, daily hassles, and negative mood, the PSA scale evaluates maternal stress that is directly related to her pregnancy including concern about the health of her baby, fear of miscarriage, and worry about the development of pregnancy-related medical problems. Accumulating evidence indicates that that measures of pregnancy-specific anxiety are better predictors of shortened gestation (Roesch et al., 2004), fetal behavior (DiPietro et al., 2002), and infant outcomes (DiPietro et al., 2006; Huizink et al., 2003) than measures of generalized maternal distress. Overall, our finding that pregnancy-specific anxiety, but not generalized distress, was associated with impairments in infant cognitive development is consistent with these studies. The variability among studies relying on self-report of generalized maternal stress (Brouwers et al., 2001; DiPietro et al., 2006; Huizink et al., 2003; Van den Bergh, 1990) may be due to fact that these measures do not capture maternal stress related to the well being of her developing fetus that is unique to pregnancy. Two published studies examined the association between pregnancy-specific anxiety and infant cognitive development and found that elevated pregnancy-specific anxiety during the third trimester predicted poorer performance on tasks assessing infant development (DiPietro et al., 2006; Huizink et al., 2003). Although the effect of timing differed between our study and the two other published studies, our findings are consistent with evidence that the fetus is more vulnerable to maternal stress early in gestation (Mueller & Bale, 2007; Schneider et al., 1999). Further, we have found that elevated pregnancy-specific anxiety early in gestation is associated with brain development in childhood (Buss et al., in press).

Prenatal Maternal Stress Is Associated With the Trajectory of Infant Development

The current findings indicate that while prenatal exposure to maternal stress and stress hormones predict infant development, significant associations do not emerge until 12 months of age. Prenatal maternal cortisol levels were, however, associated with the trajectory of infant development over the first postnatal year. Individual patterns of low maternal cortisol early and high maternal cortisol late in pregnancy predicted an accelerated trajectory of infant cognitive development over the first postnatal year, resulting in improved performance on the MDI at 12 months. It is not clear why significant differences in performance on the MDI did not emerge until 12 months of age. During the prenatal period organs and organ systems are forming and thus are especially vulnerable to both organizing and disorganizing influences. Disruption at this early stage of development has the potential to alter developmental trajectories, resulting in developmental impairments that only emerge as that capacity develops. Thus, as cognitive functions come "on line" we may begin to detect delays that were not apparent previously. Consistent with this hypothesis there is evidence from experimental studies with rodents that exposure to early life stress can have consequences for cognitive functioning that are not observable until adulthood (Brunson et al., 2005) as well as from human observational studies associating size at birth with adult health outcomes (Barker, 1998, 2002). Alternatively, it is possible that the lack of differences at 3 and 6 months may be due to limitations in our ability to evaluate children at these young ages.

Postnatal Influences on Development

There is strong evidence across a wide range of species that early postnatal experiences influence the development of individual differences (Coplan et al., 1996; Dawson & Ashman, 2000; Levine, 1957). Maternal stress, anxiety, and depression during infancy have pervasive influences on development including enhanced stress reactivity, poor emotional regulation, and impaired cognitive development (Coghill, Caplan, Alexandra, Robson, & Kumar, 1986; Dawson & Ashman, 2000; Essex, Klein, Cho, & Kalin, 2002; Kaplan, Bachorowski, Smoski, & Hudenko, 2002). It is important to note the observed effects of prenatal exposure to maternal stress on infant development were independent of postnatal maternal psychological distress (anxiety, stress, and depression) and parenting-specific stress evaluated at 3, 6, or 12 months. Furthermore, the effects of prenatal stress were observed in our full-term cohort and were not explained by sociodemographic factors including household income, maternal education, age, race, or marital status or to maternal pregnancy-related medical risk factors. Our findings that prenatal maternal stress measures were significantly associated with cognitive development in full-term infants, and that this association was not mediated or moderated by relevant postnatal factors, add strong support to the argument that prenatal experiences and exposures have a significant role in shaping developmental outcomes.

Programming Influences of Prenatal Maternal Stress

The current findings in a healthy sample of full-term infants, demonstrate that both psychological (pregnancy-specific anxiety) and endocrine (cortisol) indicators of maternal stress are associated with poorer performance on the MDI of the BSID at 1 year of age. For both psychosocial and endocrine indicators of stress, timing of exposure was critical. Optimal performance on the MDI at 12 months of age was observed in infants of mothers with low levels of cortisol early in pregnancy and high levels near term. Despite the finding that fetal exposure early in pregnancy to maternal cortisol and pregnancy-specific anxiety had similar effects on infant mental development, these two measures were not correlated and they exerted independent influences on development. This finding is consistent with the majority of published studies that have failed to discover a significant relation between maternal psychological state and maternal cortisol during pregnancy (Davis et al., 2007; McCool & Susman, 1994; Petraglia et al., 2001; de Weerth & Buitelaar, 2005). Future research will have to examine other stress responsive hormones, as well as vascular or immune pathways that could be mechanisms by which increases in maternal psychosocial stress might affect the fetus.

Limitations

First, because this study relied on naturally occurring variations in maternal cortisol, rather than experimental manipulations, it is difficult to separate the effects of cortisol from the consequences of other factors that might contribute to this association such as components of the postnatal environment including postpartum depression or maternal stress or from shared genetic effects. In the current study, however, postpartum psychological state was not associated with infant development, and adjusting for postpartum maternal psychological state did not alter the relation between prenatal maternal stress measures and infant cognitive development, supporting our conclusion that prenatal exposures were responsible for this association. Furthermore, the current findings are consistent with animal models where random assignment is possible (e.g., Catalani et al., 2002; Weinstock, 2001) as well as recent human studies that have evaluated the consequences for development of randomly occurring stressful events, such as natural disasters (Laplante et al., 2004; Yehuda et al., 2005). Second, the current study included a low-risk sample of women with healthy, full-term pregnancies. It is probable that examination of women experiencing more severe stress or who have clinically significant psychopathology would reveal more profound programming influences on developmental outcomes. Third, maternal cortisol is an indirect measure of fetal exposure. However, there is evidence with human fetuses documenting that maternal and fetal cortisol levels are significantly correlated (Gitau et al., 1998; Gitau et al., 2001).

Conclusions

Studies from animal models have demonstrated that early life environmental factors, including exposure to maternal stress and stress hormones, can cause structural and functional changes to the developing nervous system that persist throughout the lifespan. This study is among the first to evaluate the effects of endocrine and psychosocial indicators of prenatal maternal stress for infant development. Data indicate that both endocrine and pregnancy-specific psychosocial indicators of stress changed over the course of gestation and that both independently predicted infant cognitive development at 1 year of age. These effects were strongly dependent on the timing of exposure and could not be accounted for by postnatal indicators of maternal distress. Elevated levels of maternal psychosocial stress and maternal cortisol early in gestation were associated with poorer cognitive functioning at 1 year of age. In contrast, later in gestation elevated maternal cortisol was associated with improved performance on the MDI at 1 year. The current study provides new evidence that prenatal exposure to maternal cortisol and maternal psychosocial stress exerts programming effects on the fetus with persisting consequences for infant behavior. During the prenatal period, the mother's physiology directly affects the fetus and has the opportunity to shape developmental trajectories. It may be for this reason that prenatal factors exerted a stronger influence on MDI performance at 1 year as compared to more proximal maternal factors. It is our intention to continue to follow this cohort to determine if there is a persisting influence of prenatal stress on other domains of functioning including emotional regulation and whether these effects persist to later development.

References

Albidin, R. R. (1995). *Parenting Stress Index* (3rd ed.). Odessa, FL: Psychological Assessment Resources.

Austin, M. P., & Leader, L. (2000). Maternal stress and obstetric and infant outcomes: Epidemiological findings and neuroendocrine mechanisms. *Australian & New Zealand Journal of Obstetrics & Gynaecology, 40*, 331–337.

Barker, D. J. (1998). In utero programming of chronic disease. *Clinical Science, 95*, 115–128.

Barker, D. J. (2002). Fetal programming of coronary heart disease. *Trends in Endocrinology and Metabolism, 13*, 364–368.

Bayley, N. (1993). *Balyley Scales of Infant Development*. (2nd ed.). San Antonio, TX: Psychological Corporation.

Beitens, I. Z., Bayard, F., Ances, I. G., Kowarski, A., & Migeon, C. J. (1973). The metabolic clearance rate, blood production, interconversion and transplacental passage of cortisol and cortisone in pregnancy near term. *Pediatric Research, 7*, 509–519.

Bergman, K., Sarkar, P., O'Connor, T. G., Modi, N., & Glover, V. (2007). Maternal stress during pregnancy predicts cognitive ability and fearfulness in infancy. *Journal of the American Academy of Child and Adolescent Psychiatry, 46*, 1454–1463.

Brabham, T., Phelka, A., Zimmer, C., Nash, A., Lopez, J. F., & Vazquez, D. (2000). Effects of prenatal dexamethasone on spatial learning and response to stress is influenced by maternal factors. *American Journal of Physiology. Regulatory, Integrative, Comparative, and Physiology, 279*, R1899–R1909.

Brouwers, E. P. M., van Baar, A. L., & Pop, V. J. M. (2001). Maternal anxiety during pregnancy and subsequent infant development. *Infant Behavior & Development, 24*, 95–106.

Brown, R. W., Diaz, R., Robson, A. C., Kotelevtsev, Y., Mullins, J. J., Kaufman, M. H., et al. (1996). The onogeny of 11β-hydroxysteroid dehydrogenase Type 2 and mineralicorticoid receptor gene expression reveal intricate control of glucocorticoid action in development. *Endocrinology, 137*, 794–797.

Brunson, K. L., Kramár, E., Lin, B., Chen, Y., Colgin, L. L., Yanagihara, T. K., et al. (2005). Mechanisms of late-onset cognitive decline after early-life stress. *Journal of Neuroscience, 25*, 9328–9338.

Bruschettini, M., van den Hove, D. L. A., Gazzolo, D., Steinbusch, H. W. M., & Blanco, C. E. (2006). Lowering the dose of antenatal steroids: The effects of a single course of betamethasone on somatic growth and brain cell proliferation in the rat. *American Journal of Obstetrics and Gynecology, 194*, 1341–1346.

Buss, C., Davis, E. P., Muftuler, L. T., Head, K., & Sandman, C. A. (in press). High pregnancy anxiety during mid-gestation is associated with decreased gray matter density in 6-9-year-old children. *Psychoneuroendocrinology*.

Casolini, P., Cigliana, G., Alema, G. S., Ruggieri, V., Angelucci, L., & Catalani, A. (1997). Effect of increased maternal corticosterond during lactation on hippoccampal corticosteroid receptors, stress response and learning in offspring in the early stages of life. *Neuroscience, 79,* 1005–1012.

Catalani, A., Casolini, P., Cigliana, G., Scaccianoce, S., Consoli, C., Cinque, C., et al. (2002). Maternal corticosterone influences behavior, stress response and corticosteroid receptors in the female rat. *Pharmacology, Biochemistry and Behavior, 73,* 105–114.

Catalani, A., Casolini, P., Scaccianoce, S., Patacchioli, F. R., Spinozzi, P., & Angelucci, L. (2000). Maternal corticosterone during lactation permanently affects brain corticosteroid receptors, stress response and behaviour in rat progeny. *Neuroscience, 100,* 319–325.

Clarke, A. S., Wittwer, D. J., Abbott, D. H., & Schneider, M. L. (1994). Long-term effects of prenatal stress on HPA activity in juvenile rhesus monkeys. *Developmental Psychobiology, 27,* 257–269.

Coe, C. L., Kramer, M., Czeh, B., Gould, E., Reeves, A. J., Kirschbaum, C., et al. (2003). Prenatal stress diminishes neurogenesis in the dentate gyrus of juvenile rhesus monkeys. *Biological Psychiatry, 54,* 1025–1034.

Coe, C. L., & Lubach, G. R. (2005). Developmental consequences of antenatal dexamethasone treatment in nonhuman primates. *Neuroscience and Biobehavioral Review, 29,* 227–235.

Coghill, S. R., Caplan, H. L., Alexandra, H., Robson, K., & Kumar, R. (1986). Impact of maternal postnatal depression on cognitive development of young children. *British Medical Journal, 292,* 1165–1167.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior, 24,* 385–396.

Coplan, J. D., Andrews, M. W., Rosenblum, L. A., Owens, M. J., Friedman, S., Gorman, J. M., et al. (1996). Persistent elevations of cerebrospinal fluid concentrations of corticotropin-releasing factor in adult nonhuman primates exposed to early-life stressors: Implications for the pathophysiology of mood and anxiety disorders. *Proceedings of the National Academy of Sciences of the United States of America, 93,* 1619–1623.

Copper, R. L., Goldenberg, R. L., Elder, N., Swain, M., Norman, G., Ramsey, R., et al. (1996). The preterm prediction study: Maternal stress is associated with spontaneous preterm birth at less than thirty-five weeks' gestation. *American Journal of Obstetrics and Gynecology, 175,* 1286–1292.

Culhane, J. F., Rauh, V., McCollum, K. F., Hogan, V. K., K, A., & Wadhwa, P. (2001). Maternal stress is associated with bacterial vaginosis in human pregnancy. *Maternal Child Health, 5,* 127–134.

Davis, E. P., Glynn, L. M., Dunkel Schetter, C., Hobel, C., Chicz-DeMet, A., & Sandman, C. A. (2007). Prenatal exposure to maternal depression and cortisol influences infant temperament. *Journal of the American Academy of Child and Adolescent Psychiatry, 46,* 737–746.

Davis, E. P., Snidman, N., Wadhwa, P. D., Dunkel Schetter, C., Glynn, L., & Sandman, C. A. (2004). Prenatal maternal anxiety and depression predict negative behavioral reactivity in infancy. *Infancy, 6,* 319–331.

Davis, E. P., Townsend, E. L., Gunnar, M. R., Georgieff, M. K., Guiang, S. F., Cifuentes, R. F., et al. (2004). Effects of prenatal corticosteroid exposure on regulation of stress physiology in healthy premature infants. *Psychoneuroendocrinology, 29,* 1028–1036.

Davis, E. P., Townsend, E. L., Gunnar, M. R., Guiang, S. F., Lussky, R. C., Cifuentes, R. F., et al. (2006). Antenatal betamethasone treatment has a persisting influence on infant HPA axis regulation. *Journal of Perinatology, 26,* 147–153.

Dawson, G., & Ashman, S. B. (2000). On the origins of a vulnerability to depression: The influence of the early social environment on the development of psychobiological systems related to risk for affective disorders. In C. A. Nelson (Ed.), *The effects of early adversity on neurobehavioral development: The Minnesota symposium on child psychology* (Vol. 31, pp. 245–279). Mahwah, NJ: Erlbaum.

de Weerth, C., & Buitelaar, J. (2005). Physiological stress reactivity in human pregnancy—A review. *Neuroscience and Behavioral Reviews, 29,* 295–312.

de Weerth, C., van Hees, Y., & Buitelaar, J. (2003). Prenatal maternal cortisol levels and infant behavior during the first 5 months. *Early Human Development, 74,* 139–151.

Dickerson, P. A., Lally, B. E., Gunnel, E., Birkle, D. L., & Salm, A. K. (2005). Early emergence of increased fearful behavior in prenatally stressed rats. *Physiology & Behavior, 86,* 586–593.

Diorio, D., Viau, V., & Meaney, M. J. (1993). The role of the medial prefrontal cortex (cingulate gyrus) in the regulation of hypothalamic-pituitary-adrenal responses to stress. *Journal of Neuroscience, 13,* 3839–3847.

DiPietro, J. A., Ghera, M. M., Costigan, K., & Hawkins, M. (2004). Measuring the ups and downs of pregnancy stress. *Journal of Psychosomatic Obstetrics & Gynecology, 25,* 189–201.

DiPietro, J. A., Hilton, S. C., Hawkins, M., Costigan, K. A., & Pressman, E. K. (2002). Maternal stress and affect influence fetal neurobehavioral development. *Developmental Psychology, 38,* 659–668.

DiPietro, J. A., Novak, M. F. S. X., Costigan, K. A., Atela, L. D., & Ruesing, S. P. (2006). Maternal psychological distress during pregnancy in relation to child development at age two. *Child Development, 77,* 573–587.

Dole, N., Savitz, D. A., Hertz-Picciotto, I., Siega-Riz, A. M., McMahon, M. J., & Buekens, P. (2003). Maternal stress and preterm birth. *American Journal of Epidemiology, 157,* 14–24.

Drake, A. J., Tang, J. I., & Nyirenda, M. J. (2007). Mechanisms underlying the role of glucocorticoids in the early life programming of adult disease. *Clinical Science, 113*, 219–232.

Essex, M. J., Klein, M. H., Cho, E., & Kalin, N. H. (2002). Maternal stress beginning in infancy may sensitize children to later stress exposure: Effects on cortisol and behavior. *Biological Psychiatry, 52*, 776–784.

French, N., Hagan, R., Evans, S. F., Godfrey, M., & Newnham, J. (1999). Repeated antenatal corticosteroids: Size at birth and subsequent development. *American Journal of Obstetrics and Gynecology, 180*, 114–121.

Giannopoulos, G., Jackson, K., & Tulchinsky, D. (1982). Glucocorticoid metabolism in human placenta, dicidua, myometrium, and fetal membranes. *Journal of Steroid Biochemistry, 17*, 371–374.

Gitau, R., Cameron, A., Fisk, N., & Glover, V. (1998). Fetal exposure to maternal cortisol. *Lancet, 352*, 707–708.

Gitau, R., Fisk, N., Teixerira, J., Cameron, A., & Glover, V. (2001). Fetal hypothalamic-pituitary-adrenal stress responses to invasive procedures are independent of maternal responses. *Journal of Clinical Endocrinology and Metabolism, 86*, 104–109.

Gluckman, P. D., & Hanson, M. A. (2004). Living with the past: Evolution, development, and patterns of disease. *Science, 305*, 1733–1736.

Glynn, L. M., Dunkel Schetter, C., Hobel, C., & Sandman, C. A. (2008). Pattern of perceived stress and anxiety in pregnancy predict preterm birth. *Health Psychology, 27*, 42–51.

Glynn, L., Wadhwa, P. D., Dunkel Schetter, C., & Sandman, C. A. (2001). When stress happens matters: The effects of earthquake timing on stress responsivity in pregnancy. *American Journal of Obstetrics and Gynecology, 184*, 637–642.

Gutteling, B. M., de Weerth, C., & Buitelaar, J. K. (2005). Prenatal stress and children's cortisol reaction to the first day of school. *Psychoneuroendcrinology, 30*, 541–549.

Hacking, D., Watkins, A., Fraser, S., Wolfe, R., & Nolan, T. (2001). Respiratory distress syndrome and antenatal corticosteroid treatment in premature twins. *Archives of Disease in Childhood: Fetal and Neonatal Edition, 85*, F77–F78.

Hauser, J., Dettling-Artho, A., Pilloud, S., Maier, C., Feldon, J., & Pryce, C. R. (2007). Effects of prenatal dexamethasone treatment on postnatal physical, endocrine, and social development in the common marmoset monkey. *Endocrinology, 148*, 1813–1822.

Hobel, C. J. (1982). Identifying the patient at risk. In R. J. Bolognese, R. H. Schwartz, & J. Schneider (Eds.), *Perinatal medicine: Management of the high risk fetus and neonate* (pp. 3–28). Baltimore: Williams & Wilkins.

Hobel, C., & Culhane, J. (2003). Role of psychosocial and nutritional stress on poor pregnancy outcome. *Journal of Nutrition, 133*(5 Suppl. 2), 1709S–1717S.

Huizink, A. C., De Medina, P. G., Mulder, E. J., Visser, G. H., & Buitelaar, J. K. (2002). Psychological measures of prenatal stress as predictors of infant temperament. *Journal of the American Academy of Child and Adolescent Psychiatry, 41*, 1078–1085.

Huizink, A. C., Robles de Medina, P. G., Mulder, E. J., Visser, G. H., & Buitelaar, J. K. (2003). Stress during pregnancy is associated with developmental outcome in infancy. *Journal of Child Psychology and Psychiatry, 44*, 810–818.

Jacobson, L., & Sapolsky, R. (1991). The role of the hippocampus in feedback regulation of the hypothalamic pituitary adrenocortical axis. *Endocrine Reviews, 12*, 118–134.

Kaplan, P. S., Bachorowski, J., Smoski, M. J., & Hudenko, W. J. (2002). Infants of depressed mothers, although competent learners, fail to learn in response to their own mother's infant-directed speech. *Psychological Science, 13*, 268–271.

Kapoor, A., Dunn, E., Kostaki, A., Andrews, M. H., & Matthews, S. G. (2006). Fetal programming of the hypothalamic-pituitary-adrenal function: Prenatal stress and glucocorticoids. *Journal of Physiology, 572*, 31–44.

Kapoor, A., Petropoulos, S., & Matthews, S. G. (2008). Fetal programming of hypothalamic–pituitary–adrenal (HPA) axis function and behavior by synthetic glucocorticoids. *Brain Research Reviews, 57*, 586–595.

King, B. R., Nicholson, R. C., & Smith, R. (2001). Placental corticotrophin-releasing hormone, local effects and fetomaternal endocrinology. *Stress, 4*, 219–233.

Laplante, D. P., Barr, R. G., Brunet, A., Du Fort, G. G., Meaney, M. J., Saucier, J. F., et al. (2004). Stress during pregnancy affects general intellectual and language functioning in human toddlers. *Pediatric Research, 56*, 400–410.

Levine, S. (1957). Infantile experience and resistance to physiological stress. *Science, 126*, 405.

Lowy, M. T. (1994). Adrenalectomy attenuates kainic acid induced spectrin proteolysis and heat shock protein 70 in the hippocampus and cortex. *Neurochemistry, 63*, 886–894.

Lundy, B. L., Jones, N. A., Field, T., Nearing, G., Davalos, M., Pietro, P. A., et al. (1999). Prenatal depression effects on neonates. *Infant Behavior and Development, 22*, 119–129.

MacArthur, B. A., Howie, R. N., Dezoete, J. A., & Elkins, J. (1982). School progress and cognitive development of 6-year-old children whose mothers were treated antenatally with betamethasone. *Pediatrics, 70*, 99–105.

Maccari, S., & Morley-Fletcher, S. (2007). Effects of prenatal restraint stress on the hypothalamus-pituitary-adrenal axis and related behavioral and neurobiological alterations. *Psychoneuroendocrinology, 32*, S10–S15.

Marcus, S. M., Flynn, H. A., Blow, F. C., & Barry, K. L. (2003). Depressive symptoms among pregnant women screened in obstetrics settings. *Journal of Women's Health, 12*, 373–380.

Mastorakos, G., & Ilias, I. (2003). Maternal and fetal hypothalamic-pituitary-adrenal axes during pregnancy and postpartum. *Annals of the New York Academy of Science, 997*, 136–149.

Matthews, S. G. (2000). Antenatal glucocorticoids and programming of the developing CNS. *Pediatric Research, 47,* 291–300.

McCool, W. F., & Susman, E. J. (1994). Cortisol reactivity and self report anxiety in the antepartum: Predictors of maternal intrapartal outcomes in gravid adolescents. *Journal of Psychosomatic Obstetrics and Gynecology, 15,* 9–18.

Mennes, M., Stiers, P., Lagae, L., & Van den Bergh, B. (2006). Long term cognitive sequelae of antenatal maternal anxiety: Involvement of the orbitofrontal cortex. *Neuroscience and Biobehavioral Reviews, 30,* 1078–1086.

Morley, R., Blair, E., Dwyer, T., & Owens, J. (2002). Is birthweight a good marker for gestational exposures that increase the risk of adult disease? *Pediatric and Perinatal Epidemiology, 16,* 194–199.

Mueller, B. R., & Bale, T. L. (2007). Early prenatal stress impact on coping strategies and learning performance is sex dependent. *Physiology and Behavior, 91,* 55–65.

Munck, A., Guyre, P. M., & Holbrook, N. J. (1984). Physiological functions of glucocorticoids in stress and their relation to pharmacological actions. *Endocrine Reviews, 5,* 25–44.

Murphy, V. E., Smith, R., Giles, W. B., & Clifton, V. L. (2006). Endocrine regulation of human fetal growth: The role of the mother, placenta, and fetus. *Endocrine Reviews, 27,* 141–169.

O'Connor, T. G., Heron, J., & Glover, V. (2002). Antenatal anxiety predicts child behavioral/emotional problems independently of postnatal depression. *Journal of the American Academy of Child and Adolescent Psychiatry, 41,* 1470–1477.

Petraglia, F., Florio, P., Nappi, C., & Genazzani, A. R. (1996). Peptide signaling in human placenta and membranes: Autocrine, paracrine, and endocrine mechanisms. *Endocrine Review, 17,* 156–186.

Petraglia, F., Hatch, M. C., Lapinski, R., Stomati, M., Reis, F. M., Cobellis, L., et al. (2001). Lack of effect of psychosocial stress on maternal corticotropin-releasing factor and catecholamine levels at 28 weeks' gestation. *Journal of the Society for Gynecologic Investigation, 8,* 83–88.

Power, M. L., Bowman, M. E., Smith, R., Ziegler, T. E., Layne, D. G., Schulkin, J., et al. (2006). Pattern of maternal serum corticotropin-releasing hormone concentration during pregnancy in the common marmoset (*Callithrix jacchus*). *American Journal of Primatology, 68,* 181–188.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Application and data analysis methods* (Vol. 1). Thousand Oaks, CA: Sage.

Rini, C. K., Dunkel-Schetter, C., Wadhwa, P. D., & Sandman, C. A. (1999). Psychological adaptation and birth outcomes: The role of personal resources, stress, and sociocultural context in pregnancy. *Health Psychology, 18,* 333–345.

Roesch, S. C., Dunkel-Schetter, C., Woo, G., & Hobel, C. J. (2004). Modeling the types and timing of stress in pregnancy. *Anxiety, Stress & Coping, 17,* 87–102.

Sanchez, M. M., Young, L. J., Plotsky, P. M., & Insel, T. R. (2000). Distribution of corticosteroid receptors in the rhesus brain: Relative absence of glucocorticoid receptors in the hippocampal formation. *Journal of Neuroscience, 20,* 4657–4668.

Sandman, C. A., Glynn, L. M., Dunkel Schetter, C., Wadhwa, P. D., Garite, T., & Hobel, C. (2006). Elevated maternal cortisol early in pregnancy predicts third trimester levels of placental corticotropin releasing hormone (CRH): Priming the placental clock. *Peptides, 27,* 1457–1463.

Santor, D. A., & Coyne, J. C. (1997). Shortening the CES-D to improve its ability to detect cases of depression. *Psychological Assessment, 9,* 233–243.

Scaccianoce, S., Catalani, A., Lombardo, K., Consoli, C., & Angelucci, L. (2001). Maternal glucocorticoid hormone influences nerve growth factor expression in the developing rat brain. *Developmental Neuroscience, 12,* 2881–2884.

Schneider, M. L. (1992). Prenatal stress exposure alters postnatal behavioral expression under conditions of novelty challenge in rhesus monkey infants. *Developmental Psychobiology, 25,* 529–540.

Schneider, M. L., & Coe, C. L. (1993). Repeated social stress during pregnancy impairs neuromotor development of the primate infant. *Journal of Developmental and Behavioral Pediatrics, 14,* 81–87.

Schneider, M. L., Coe, C. L., & Lubach, G. R. (1992). Endocrine activation mimics the adverse effects of prenatal stress on the neuromotor development of the infant primate. *Developmental Psychobiology, 25,* 427–439.

Schneider, M. L., Roughton, E. C., Koehler, A. J., & Lubach, G. R. (1999). Growth and development following prenatal stress exposure in primates: An examination of ontogenetic vulnerability. *Child Development, 70,* 263–274.

Singer, J., & Willett, J. (Eds.). (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* New York: Oxford University Press.

Speilberger, C. (1983). *Manual for the State-Trait Anxiety Inventory.* Palo Alto, CA: Consulting Psychologists Press.

Spielberger, C. (1983). *State-Trait Anxiety Inventory.* Redwood City, CA: Mind Garden.

Spinillo, A., Viazzo, F., Colleoni, R., Chiara, A., Cerbo, R. M., & Fazzi, E. (2004). Two-year infant neurodevelopmental outcome after single or multiple antenatal courses of corticosteroids to prevent complications of prematurity. *American Journal of Obstetrics and Gynecology, 191,* 217–224.

Trejo, J. L., Cuchillo, I., Machin, C., & Rua, C. (2000). Maternal adrenalectomy at the early onset of gestation impairs the postnatal development of the rat hippocampal formation: Effects on cell numbers and differentiation, connectivity and calbindin-D28K immunoreactivity. *Journal of Neuroscience Research, 62,* 644–667.

Uno, H., Eisele, S., Sakai, A., Shelton, S., Baker, E., DeJesus, O., et al. (1994). Neurotoxicity of glucocorticoids in the primate brain. *Hormone and Behavior, 28,* 336–348.

Uno, H., Lohmiller, L., Thieme, C., Kemnitz, J. W., Engle, M. J., Roecker, E. B., et al. (1990). Brain damage induced by prenatal exposure to dexamethasone in fetal macaques. I. Hippocampus. *Brain Research Developmental Brain Research, 53,* 157–167.

Van den Bergh, B. (1990). The influence of maternal emotion during pregnancy on fetal and neonatal behavior. *Pre- and Peri-natal Psychology Journal, 5,* 119–130.

Van den Hove, D. L. A., Blanco, C. E., Aendekerk, B., Desbonnet, L., Bruschettini, M., Steinbusch, H. P., et al. (2005). Prenatal restraint stress and long-term affective consequences. *Developmental Neuroscience, 27,* 313–320.

Wapner, R. J., Sorokin, Y., Mele, L., Johnson, F., Dudley, D. J., Spong, C. Y., et al. (2007). Long-term outcomes after repeat doses of antenatal corticosteroids. *New England Journal of Medicine, 357,* 1190–1198.

Weinstock, M. (2001). Alterations induced by gestational stress in brain morphology and behaviour of the offspring. *Progress in Neurobiology, 65,* 427–451.

Welberg, L. A., & Seckl, J. (2001). Prenatal stress, glucocorticoids and the programming of the brain. *Journal of Neuroendocrinology, 13,* 113–128.

Welberg, L. A., Seckl, J. R., & Holmes, M. C. (2001). Prenatal glucocorticoid programming of brain corticosteroid receptors and corticotrophin-releasing hormone: Possible implications for behaviour. *Neuroscience, 104,* 71–79.

Yehuda, R., Engel, S. M., Brand, S. R., Seckle, J., Marcus, S. M., & Berkowitz, G. S. (2005). Transgenerational effects of posttraumatic stress disorder in babies of mothers exposed to the World Trade Center attacks during pregnancy. *Journal of Clinical Endocrinology & Metabolism, 90,* 4115–4118.

Zarrow, M. X., Philpott, J. E., & Denenberg, V. H. (1970). Passage of 14-C-4 corticosterone from the rat mother to the fetus and neonate. *Nature, 226,* 1058–1059.

Child Development, January/February 2010, Volume 81, Number 1, Pages 149–165

Mothers' Antenatal Depression and Their Children's Antisocial Outcomes

Dale F. Hay
Cardiff University

Susan Pawlby
King's College London

Cerith S. Waters and Oliver Perra
Cardiff University

Deborah Sharp
University of Bristol

Interviews of 120 British adolescents and their parents (80% of a random sample of antenatal patients drawn from a representative urban population and followed longitudinally) revealed that 40 (33%) had been arrested and/or had a diagnosis of *DSM–IV* conduct disorder by 16 years of age; of those, 18 (45%) had committed violent acts. Depression in pregnancy significantly predicted violence in adolescence, even after controlling for the family environment, the child's later exposure to maternal depression, the mother's smoking and drinking during pregnancy, and parents' antisocial behavior. Mothers with a history of conduct problems were at elevated risk to become depressed in pregnancy, and the offspring of depressed women had a greater chance of becoming violent by age 16.

The aim of our study was to examine links between exposure to maternal depression in pregnancy and antisocial outcomes in a longitudinal sample of British children who were studied at intervals until the age of 16. It has been shown that the mother's mental health problems in pregnancy predict disruptive behavior disorders in offspring. In particular, the mother's symptoms of anxiety predict behavioral problems (O'Connor, Heron, Golding, Beveridge, & Glover, 2002) and attention deficit hyperactivity disorder (Van den Bergh & Marcoen, 2004; Van den Bergh, Mennes, et al., 2005; Van den Bergh, Mulder, Mennes, & Glover, 2005; Van den Bergh et al., 2006). In a large Finnish cohort, mothers' reports of depression during pregnancy predicted antisocial behavior in their adult offspring 30 years later (Maki et al., 2003). In this study we concentrate attention on the more prevalent clinical disorder, depression, while simultaneously controlling for the level of mothers' anxiety symptoms in pregnancy.

Although these studies have found evidence for an association between mothers' mental health problems in pregnancy and children's outcomes, it is not clear whether that association reflects a direct effect of the mother's mental state on the fetus or whether it is accounted for by other social and biological processes. Several plausible hypotheses can be generated to account for these links between antenatal exposure to maternal depression and antisocial outcomes. The association could be explained by: (a) *direct effects on the developing fetus* from the biological correlates of the mothers' symptoms; (b) *the association of depression in pregnancy with more general measures of environmental adversity*, which are well-known predictors of antisocial behavior; (c) *re-exposure to the maternal depression*, whereby infants who are exposed in pregnancy are at elevated risk for re-exposure and consequent disruption of the postnatal environment; (d) *indirect effects of antenatal depression on the developing fetus* mediated by mothers' smoking, drinking, and drug taking in pregnancy; and (e) *parents' antisocial behavior*, as women who experience depression in pregnancy may also be at genetic risk for antisocial behavior and may transmit both risks to offspring. There is evidence bearing on each of these hypotheses.

The SLCDS has been funded by the Medical Research Council UK Project Grants G89292999N and G9539876N awarded to the late Channi Kumar, Deborah Sharp, and Dale Hay; the Psychiatry Research Trust; and the South West G.P. Trust. The present data coding and analyses were partially supported by an Economic and Social Research Council studentship to Cerith S. Waters and Medical Research Council UK Programme Grant GO400086. We thank Helen Allen for her contribution to the retention of the sample, and the families who participated in the study for 16 years.

Correspondence concerning this article should be addressed to Dale F. Hay, School of Psychology, Cardiff University, Park Place, Cardiff, Wales, UK CF10 3AT. Electronic mail may be sent to haydf@cardiff.ac.uk.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8101-0010

Evidence for the Alternative Hypotheses

Hypothesis 1: The mother's depression in pregnancy has direct effects on the developing fetus. Conduct problems, especially those that emerge in early childhood, are linked to neurobiological deficits (e.g., Raine et al., 2005), which are in turn associated with risk factors in the antenatal environment (van Goozen, Fairchild, Snoek, & Harold, 2007; Raine, 2002). Possible mechanisms whereby mothers' symptoms of mental illness affect their children include disruption of normal development of the hypothalamic-pituitary-adrenal (HPA) axis (e.g., Lundy et al., 1999; Talge et al., 2007; van Goozen et al., 2007) and effects of gestational stress on fetal growth (e.g., Borders, Grobman, Amsden, & Holl, 2007; Rahman, Bunn, Lovel, & Creed, 2007). Our aim in this study was not to test specific hypotheses about potential mechanisms but merely to establish whether there was evidence for an association with antenatal depression, once the alternative hypotheses had been tested.

Hypothesis 2: The link between antenatal symptoms and antisocial behavior is explained by preexisting environmental adversity. Antisocial behavior flourishes in adverse social environments. For example, children's aggression has been linked to socioeconomic adversity (e.g., Keenan & Shaw, 1994), exposure to adult maltreatment (e.g., Dodge, Bates, & Pettit, 1990), and high levels of violence in the community (e.g., Kupersmidt, Griesler, DeRosier, Patterson, & Davis, 1995). One well-known marker of general environmental adversity is teenage parenthood, which predicts antisocial behavior (Jaffee, Caspi, Moffitt, Belsky, & Silva, 2001) and is also associated with the mother's marital status and her history of depressive illness (Romano, Zoccolillo, & Paquette, 2006). It is possible that the effects of antenatal depression disappear when environmental adversity is taken into account.

Hypothesis 3: The link between antenatal depression and antisocial outcomes is explained by re-exposure to maternal depression. Developmental theorists have long argued that any test of early experience effects —that is, analyses of the impact of deprivation and stimulation before and immediately after birth— must be conducted with reference to individuals' experiences later in life, and the present circumstances under which they are being assessed (Cairns, 1979; Rutter, 1972; Solomon & Lessac, 1968). Thus, when trying to explain why antenatal depression predicts disruptive behavior in offspring, it is important to note that depression is a lifelong, episodic illness, and the child is likely to be re-exposed. Exposure to maternal depression during fetal life predicts later exposure postpartum (O'Hara, 1997), and postpartum depression predicts subsequent episodes in later childhood and adolescence (Halligan, Murray, Martins, & Cooper, 2006).

The mother's chronic depressive illness has been linked to children's emotional and behavioral problems (Campbell & Cohn, 1997). Maternal depression during childhood is a strong predictor of children's conduct symptoms, even when the mother's antisocial behavior is taken into account (Kim-Cohen, Moffitt, Taylor, Pawlby, & Caspi, 2005). It is therefore possible that any effects of antenatal depression are explained by the child's later exposure to the mother's illness.

Hypothesis 4: The mother's illness has indirect effects on the developing fetus that are mediated by her substance use in pregnancy. Smoking, alcohol, and drug use in pregnancy are associated with deleterious outcomes for the fetus and the child (e.g., Carvalho De Moraes Barros, et al., 2006; Mathews, Menaker, & MacDorman, 2003; Maughan, Taylor, Taylor, Butler, & Bynner, 2001). To the extent that depressed women are more likely to drink, smoke, and take legal and illegal drugs in pregnancy, their offspring are more likely to be exposed to these antenatal insults, which are thought to exert direct influence on the developing fetus.

Hypothesis 5: Many children whose mothers experience depression in pregnancy are already at familial risk for antisocial behavior. There is considerable evidence for genetic as well as environmental transmission of risk for antisocial behavior from parent to child (Moffitt, 2005), and so it is important to consider whether the association between antenatal depression and children's antisocial outcomes can be explained by parents' own antisocial tendencies. Two important issues to consider are the *co-occurrence* of the mother's depressive illness and her history of conduct disorder (CD), and *assortative mating* between antisocial women and men.

With respect to the first issue, women with an antisocial history are more likely than other mothers to expose their offspring to depression in pregnancy, because CD is often associated with depression (Romano et al., 2006; Zoccolillo, Meyers, & Assiter, 1997). Girls who have shown conduct problems in childhood are likely to experience depression in adolescence and adulthood (Rutter, Giller, & Hagell, 1998) and are at special risk for depression during pregnancy and the months after childbirth (Cassidy, Zoccolillo, & Hughes, 1996; Romano et al., 2006). Thus, a substantial proportion of women who experience depression during pregnancy have a past history of antisocial behavior.

Second, with respect to the issue of assortative mating, the offspring of women with a history of CD and current depression may also be at genetic risk through processes of paternal transmission. Antisocial women may often have antisocial partners (Galbaud du Fort, Boothroyd, Bland, Newman, & Kakuma, 2002; Krueger, Moffitt, Caspi, Bleske, & Silva, 1998), and the presence of paternal as well as maternal antisocial tendencies would increase the risk for the child. Thus, it is important to control for antisocial behavior in both biological parents.

Aim of the Study: Charting Cumulative Risk for Antisocial Behavior From Pregnancy to 16 Years

In highlighting these alternative explanations of a link between the mother's depression in pregnancy and her child's antisocial behavior, it immediately becomes clear that the different processes under investigation are not mutually exclusive. Indeed, the evidence shows that the children of mothers who are depressed in pregnancy may also be exposed to a cumulative sequence of risk factors, including the pregnant woman's anxiety symptoms, her use of substances during pregnancy, and her subsequent episodes of depressive illness. These cumulative risks associated with antenatal depression would be further exacerbated by her antisocial tendencies, the father's antisocial behavior, and general environmental adversity throughout the child's lifetime. To determine whether depression in pregnancy has direct effects, or is explained by the other risks that accumulate over the child's life, more detailed descriptive information is needed, preferably from a longitudinal sample that was first assessed in pregnancy.

Therefore, the aim of our study was to describe such accumulation of risks for antisocial behavior from the mother's pregnancy to 16 years of age in a population of British children who reside in relatively disadvantaged urban communities and are therefore already at general environmental risk for antisocial behavior, psychopathology, and poor educational and occupational outcomes. The South London Child Development Study (SLCDS) is an ongoing longitudinal study of families in two areas of South London. The original purpose of the study was to identify the environmental predictors of women's physical and mental health in pregnancy and the first year postpartum (Sharp, 1993). Subsequent follow-up assessments at 4, 11, and 16 years postpartum have tested whether the mother's perinatal illness increased the children's risk for cognitive, emotional, and behavioral problems (e.g., Hay et al., 2001, Hay, Pawlby, Angold, Harold, & Sharp, 2003; Pawlby, Hay, Sharp, Waters, & O'Keane, 2009). The SLCDS thus provides an opportunity to chart the pathway of cumulative risk from fetal life to adolescence, in a population that has been exposed to many of the community influences that promote antisocial behavior.

Within these urban communities, some families (approximately 10%) are well educated and affluent, but most are working-class families who have experienced some economic disadvantage. Examination of cumulative risk in this urban sample controls for variation in environmental context (e.g., urban–rural differences) and focuses attention on the psychological and behavioral processes that explain why some but certainly not all of the children in these communities follow pathways toward antisocial behavior.

Prediction of Violent Versus Nonviolent Antisocial Behavior

We aim to determine whether antenatal depression has different effects on different domains of antisocial behavior. It is possible that a combination of antenatal risk factors and periodically disrupted postnatal environments contribute to the development of violence in particular. Loeber et al. (1993) proposed that there is a distinct developmental pathway toward overt physical aggression and violence, as distinguished from covert antisocial acts or nonviolent conflict with authority figures. Serious violence peaks in adolescence and young adulthood, but individual differences in violent tendencies may nonetheless be influenced by earlier risk factors. Thus, in the present study, we examine the effects of depression in pregnancy on violent and nonviolent antisocial behavior in adolescence while controlling for general environmental adversity, cumulative exposure to maternal mental illness, substance use in pregnancy, and parents' antisocial behavior.

Method

The Families

A random sample of 178 women was drawn from records of consecutive antenatal patients who attended two National Health Service (NHS) general medical practices in South London in 1986 (Sharp, 1993). In the United Kingdom, 96% of the population is registered with the NHS (Hannaford, Smith, & Eliot, 2006). Only 0.5% of births in the United Kingdom take place in private hospitals (NHS, 2004/2005). Therefore, virtually all pregnant women in the two geographical catchment areas were registered for antenatal care at these two NHS medical practices, and this sampling frame permitted recruitment of a truly random sample from a representative population of infants born in these areas during 1 year of recruitment.

Of the 178 women in the random sample at the beginning of their pregnancies, 9 suffered miscarriages, intrauterine death, or stillbirth; 7 moved out of the area and at that point were excluded from the sample; 1 infant was taken into the care of social services and researchers' access was denied; and 10 declined to participate further in the study. Thus, the follow-up study of the children was based on the remaining $N = 151$ cases (89% of the random sample). At 16 years, 137 of the 151 families (91%) participated in the study. This study focuses on 120 families (80% of 151 assessed postpartum) for whom complete information was available on all variables needed to test the study hypotheses.

Characteristics of the participants with complete data at age 16 are summarized in Table 1. Compared to national norms, the sample had a higher proportion of working-class families and families from ethnic minorities (UK Office for National Statistics, 2006). Based on the 2001–2002 data collected throughout England, the deprivation scores of the two South London communities from which the random sample was drawn ranked on the 6th and 11th percentiles, where a lower rank indicates greater deprivation. The sample of $N = 120$ cases with complete data did not differ significantly from the cases assessed postpartum ($N = 151$) with respect to social class, mothers' marital status at birth, mother's age at first birth, parents' ethnic origin, the infant's birth weight, and the proportion of firstborns.

All phases of the study were approved by the Ethics Committee of the Institute of Psychiatry, King's College London. After complete description of all procedures, written informed consent was obtained from mothers and at 16 years from their adolescent offspring.

Procedure

Pregnancy and the first postnatal year. Two doctors, one from each practice, interviewed the mothers, who were not their patients, twice during pregnancy, between 14 and 20 weeks and at 36 weeks, and twice during the first postnatal year, at 3 and at 12 months.

Table 1
Characteristics of the Sample at 16 Years Postpartum (N = 120)

Variable	Description
Mother's age at birth	$M = 26.7$, $SD = 5.0$ (range = 17–44)
Mother's age at first birth	$M = 23.8$, $SD = 4.6$ (range = 17–37)
Marital status at birth	65% married, 28% cohabiting, 7% single
Social class	88% working class
Maternal education	72% basic qualifications or higher
Marital status at age 16	59% married (87% to biological father) 13% cohabiting (33% to biological father) 28% single parent household
Parent in household at 16	54% two biological parents 39% biological mother 5% biological father 2% no biological parent (living independently, $N = 1$; under supervision of Social Services, $N = 1$)
Child's sex	55% female
Child's birth order	48% firstborn
Child's ethnicity	72% White British, 6% White, non-British, 22% Other (Caribbean, African, South Asian, East Asian, mixed)
Child's age at assessment	$M = 16.3$ years, $SD = 0.2$ (range = 16–17.3)

Note. *N*s vary slightly (117–120) because of missing data on some variables.

Fourth, 11th, and 16th birthdays. Families were visited at home at 4, 11, and 16 years postpartum, and at 11 years the children were seen at school. At each time point mothers provided sociodemographic information, current and retrospective to the last visit, to interviewers who were unaware of the information collected at previous visits. Parents (in most cases, the biological mother) and children were interviewed at 11 and 16 by different researchers.

Measures

Maternal psychopathology. At both points in pregnancy and at 3 and 12 months postpartum, an assessment was made of the women's current mental state, using the Clinical Interview Schedule (CIS) to generate ICD-9 diagnoses of depression and anxiety (Goldberg, Cooper, Eastwood, Kedward, & Shepherd, 1970). The overall agreement of the reported symptoms on the CIS from the tape-recorded interviews, given as a weighted kappa coefficient, was .80. Mothers completed the Leeds

Anxiety Scale (Snaith, Bridge, & Hamilton, 1971) twice, in the second and third trimesters of pregnancy. The Leeds scale scores at each time point were averaged to provide a composite measure of symptoms of anxiety in pregnancy.

At 4, 11, and 16 years postpartum, mothers were interviewed about their current mental states and experience of symptoms retrospective to the last assessment point, using the Schedule of Affective Disorders and Schizophrenia (SADS–L; Spitzer, Endicott, & Robins, 1978). Diagnoses were made in consultation with the lead psychiatrist on the team, using Research Diagnostic Criteria (RDC).

Depression in pregnancy was defined by the combined ICD diagnoses from the second and third trimester interviews; if the woman had met ICD criteria for depression at either time, she was judged to have experienced antenatal depression. Postpartum depression was defined by the ICD diagnosis of depression at 3 months postpartum. ICD criteria were also used to diagnose depression at 12 months postpartum.

The current and "retrospective to last visit" data were used to assess the occurrence of maternal depression over the child's lifetime. Two additional variables were created to measure the children's exposure to maternal depression in early childhood (between 1 and 4 years of age) and later childhood (between 4 and 11 years of age). An ordinal variable of the number of time periods in which women were diagnosed with depression was then computed.

Sociodemographic risk factors. Characteristics of the family environment were assessed by interview at all time points (see Table 1). Dichotomous measures of risk factors in the family environment were then computed: (a) mother's *social class*, based on occupation (Goldthorpe & Hope, 1974), dichotomized into working versus middle or upper class; (b) *maternal education*, dichotomized into no qualifications versus basic qualifications or higher; (c) self-reported *cultural background*, dichotomized into White British versus other ethnic identity; (d) *mother's age at the birth of the index child*, dichotomized as teenager (19 years of age and under) versus older; (e) *mother's age at the birth of her first child*, dichotomized as teenager (19 years of age and under) versus older; (f) *mother's marital status* at the child's birth (married vs. cohabiting or single/divorced/widowed); and (g) *two-parent family structure* at the child's birth (married/cohabiting vs. single/divorced/widowed).

The mother's intellectual ability. During the home visit at 4 years, four subtests (vocabulary, arithmetic, similarities, and block design) of the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1981) were administered to the mothers and scored to a formula that takes into account correlation with the full WAIS in terms of the subtests' reliability and age-scaled scores (Brooker & Cyr, 1986). The subtests were used again at the 11- or 16-year assessment to assess those mothers who had not been seen since the first postnatal year.

Measures of mothers' smoking and drinking during pregnancy. During the second- and third-trimester interviews, mothers reported the number of cigarettes smoked per day and units of alcohol drunk per week. Composite measures of fetal exposure to cigarette smoking and alcohol were constructed by averaging the scores reported at each interview.

Mothers' antisocial behavior. Mothers' history of arrest was obtained from their interviews at 4, 11, and 16 years. At the 16-year visit, we obtained mothers' retrospective reports of their own conduct symptoms before the age of 15, using the antisocial personality disorder section of the SADS–L. All women were asked about juvenile conduct symptoms, whether or not they screened in as possible cases of antisocial personality disorder. A 10-item scale was constructed, summing women's reports of truancy, expulsion from school, rule breaking, stealing, lying, running away, vandalism, underage alcohol use, underage sex, and juvenile arrest (Cronbach's $\alpha = .75$).

Biological fathers' history of arrest. Mothers reported on the biological fathers' history of arrest at the 11- and/or 16-year interviews. Previous research has demonstrated the reliability of mothers' reports about their partners' antisocial behavior (Caspi et al., 2001). In five cases where the mother did not have custody of the child, the father was interviewed and reported on his own history of arrest.

Adolescents' antisocial behavior. At 11 and 16, parents and children were interviewed separately about the child's symptoms of clinical disorder, using the Child and Adolescent Psychiatric Assessment (CAPA; Angold et al., 1995). The CAPA is a psychiatric interview that elicits information about symptoms contributing to a wide range of *DSM–IV* (American Psychiatric Association, 1994) diagnoses. A 3-month "primary period" is used to ensure more accurate recall (Angold, Erkanli, Costello, & Rutter, 1996).

At 11 and 16 years, *DSM–IV* diagnoses and symptom scales were generated by computer algorithms, based on "combined reports," where a symptom is regarded as being present if either parent or child reports it. Diagnoses were made with reference to the functional impairment or incapacities section of the CAPA, which relates the symptoms to the adolescent's ability to function at a developmentally appropriate level in relationships with family, peers, and teachers, and in activities at school, home, and in the community. Diagnoses made using CAPA algorithms show acceptable levels of test–retest reliability (Angold & Costello, 2000).

For the purposes of this study, to identify youngsters with a history of antisocial activity, we focus on the *DSM–IV* diagnoses of CD based on parents' and adolescents' joint reports at ages 11 and 16 and the parents' and/or adolescents' reports of police contact up to the age of 16. A dichotomous measure of antisocial behavior was constructed, with adolescents judged to show antisocial tendencies if they had been arrested up to the age of 16 and/or been diagnosed with CD at age 11 and/or 16.

Violent antisocial behavior was defined in terms of the *DSM–IV* categorization of aggressive conduct symptoms (fighting, bullying, cruelty to people, theft with confrontation, and use of a weapon) and/or arrest for violent crimes (theft from a person, violent disorder, fighting, carrying an offensive weapon, assault, and actual bodily harm/grievous bodily harm). Antisocial adolescents were further categorized as showing violent behavior if they had shown at least one aggressive conduct symptom (in the context of the *DSM–IV* diagnosis of CD) and/or had been arrested for a violent crime.

Data Analysis

All variables were examined for psychometric properties and analyses appropriate to the scale of measurement were used. We first present univariate associations between depression in pregnancy and other risk factors, and between antisocial behavior and all risk factors (see Tables 3 and 4). Logistic regression models were then used to test the five alternative hypotheses regarding the association between depression in pregnancy and antisocial outcomes. We then tested whether depression in pregnancy predicted violent behavior in particular (see Table 5, Figure 1a–c). Subsequent Markov chain models were employed, using LEM 1.0 (Vermunt, 1997), to elucidate further the additive contribution of multiple risk factors to the prediction of violent behavior.

Results

Antisocial Outcomes

Incidence of arrests and DSM–IV *CD*. Forty of the 120 adolescents (33%) met criteria for clinically and socially significant antisocial behavior (see Table 2). Of the 40, 32 (88.9%) had been arrested before the age of 16 for a variety of violent and nonviolent offences (Table 2). Twenty (50%) had been diagnosed with CD at 11 and/or 16, 12 of whom had also been arrested. Adolescents who had been diagnosed with CD were significantly more likely than other children in the sample to be arrested, $\chi^2(1) = 13.64$, $p < .001$, odds ratio (OR) = 6.00 (95% confidence interval [CI] 2.16–16.64).

Of the 32 adolescents who had been arrested by age 16, 12 had committed offences involving violence against people: theft from a person, violent disorder, fighting, carrying an offensive weapon, assault, and actual bodily harm/grievous bodily harm (see Table 2). Of the 8 additional adolescents who had not been arrested, but who met *DSM–IV* criteria for CD, 6 had shown aggressive conduct symptoms. Thus, 18 adolescents were classified as having shown some evidence of violent antisocial behavior (15% of the sample, and 45% of those who were classified as antisocial at age 16).

Tests for sex differences in antisocial behavior. Twenty-two of 54 boys (40.7%) and 18 of 66 girls (27.3%) met criteria for antisocial behavior in adolescence, having been arrested or diagnosed with CD, using *DSM–IV* criteria. The sex difference was not significant. Further analyses revealed no significant sex differences in any measure of antisocial behavior. Sixteen boys (29.6%) and 16 girls (24.2%) had been arrested; 12 boys (22.2%) and 8 girls (12.1%) were diagnosed with CD at either 11 or 16. Eleven boys (20.4%) and 7 girls (10.6%) had engaged in violent antisocial behavior. In view of the absence of significant sex differences, subsequent analyses were undertaken on the whole sample.

Mothers' Experience of Mental Health Problems in Pregnancy

Thirty-eight women (31%) met ICD-9 diagnostic criteria for depression and 4 for an anxiety disorder during pregnancy, with 2 women being comorbid. Our analyses therefore focus on diagnostic caseness for the more prevalent disorder, depression, while controlling for anxiety symptoms as a continuous measure in the sample as a whole. Depressed women had significantly higher levels of anxiety symptoms in pregnancy, $M = 6.3$ ($SD = 3.4$) for depressed women versus $M = 4.3$ ($SD = 2.9$) for women who were not depressed, $t(118) = -3.27$, $p < .001$. The associations between depression in pregnancy and the other risk factors tested are summarized in Table 3.

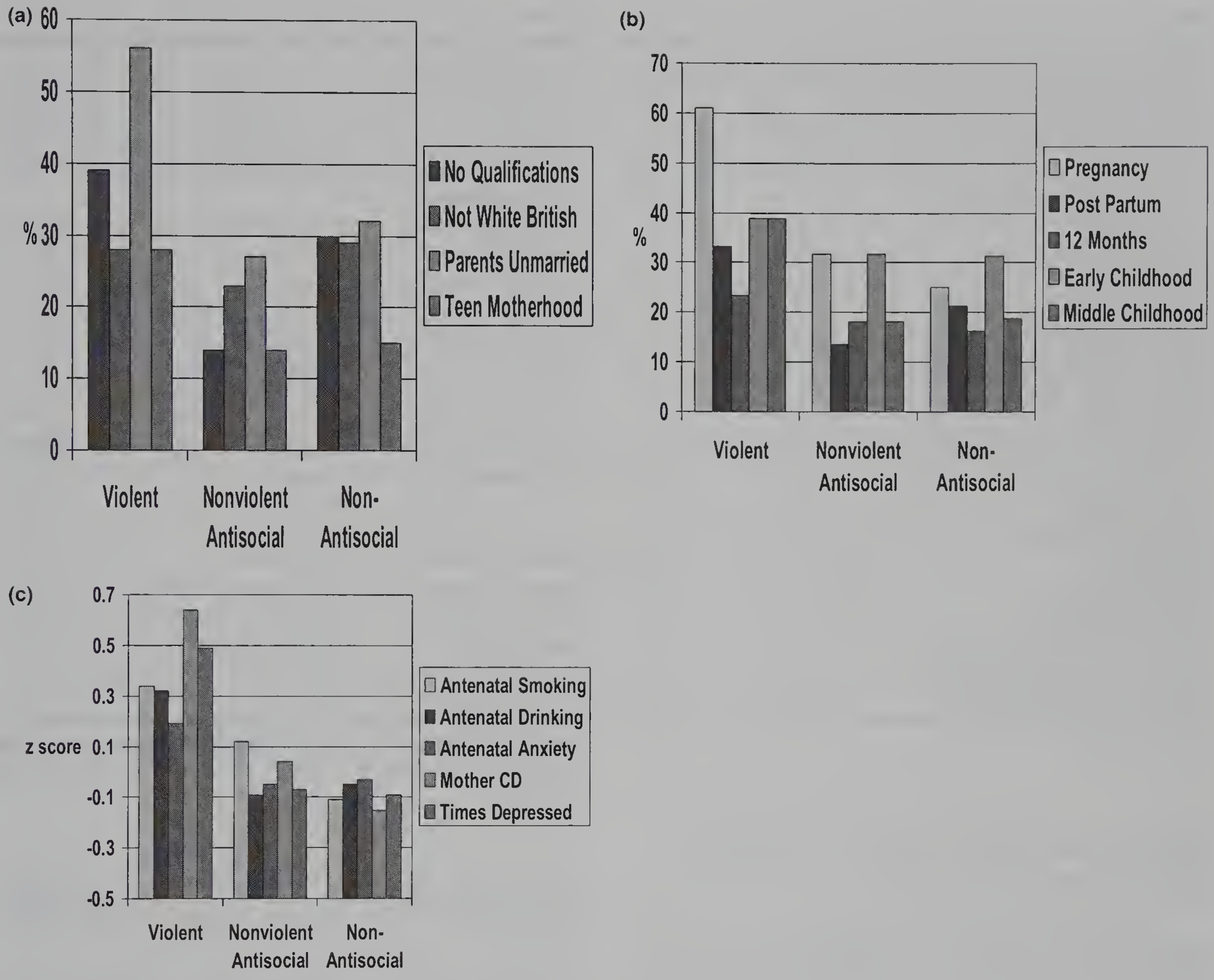


Figure 1. Prior environmental risk factors (a), exposure to maternal depression (b), and maternal risk factors (c) for violent, nonviolent antisocial, and non-antisocial adolescents.

Hypothesis 1: Antenatal Depression Has Direct Effects on Antisocial Behavior in Adolescence

The mother's ICD-9 caseness for depression in pregnancy was a significant predictor of antisocial behavior in adolescence. If mothers had not been depressed during pregnancy, 22 of 82 adolescent offspring (26.8%) showed antisocial behavior. In contrast, if mothers had been depressed in pregnancy, 18 of 38 adolescent offspring (47.4%) showed antisocial behavior, $\chi^2_{(1)} = 4.93$, $p < .03$, OR = 2.46 (95% CI 1.10–5.48). Furthermore, adolescents whose mothers had been depressed in pregnancy were at a fourfold risk for violence. When mothers had not been depressed in pregnancy, 7 of 82 adolescents (8.5%) had shown violence, as opposed to 11 of 38 adolescents (28.9%) whose mothers had been depressed at that time (see Table 5), $\chi^2(1) = 8.48$, $p < .01$, OR = 4.36 (95% CI 1.54–12.41).

Antisocial behavior in adolescence was not significantly associated with the mothers' reports of anxiety symptoms during pregnancy (see Table 4). There was no evidence for a timing effect. Anxiety symptoms in midpregnancy and in late pregnancy were both unrelated to antisocial behavior in adolescence. A logistic regression analysis revealed that the association between depression in pregnancy and antisocial behavior remained significant when controlling for anxiety symptoms in pregnancy, Wald statistic = 4.60, $p < .03$, OR = 2.51 (95% CI 1.08–5.83), as did the association between depression in pregnancy and violent behavior, Wald statistic = 6.91, $p < .01$, OR = 4.33 (95% CI 1.45–12.92).

Table 2

The Nature of Antisocial Activities Shown by the Adolescents (DSM–IV Conduct Disorder at 16 and/or Worst Offence for Which Arrested by 16; N = 40/120)

	N	%
DSM–IV conduct disorder at 11 or 16, no arrest	8	6.7
Shoplifting	6	5.0
Theft of property	1	0.8
Truancy	2	1.7
Fraud/intention to defraud	1	0.8
Rule breaking/underage drinking	1	0.8
Drug dealing/possession	1	0.8
Trespassing	2	1.7
Vandalism/damage to property	3	2.5
Joy riding	1	0.8
Fire setting/arson	1	0.8
Breaking and entering	1	0.8
Fighting[a]	6	5.0
Violent disorder	1	0.8
Theft from a person[a]	1	0.8
Possession of an offensive weapon[a]	2	1.7
Assault on a person (no injury)[a]	1	0.8
Actual/grievous bodily harm[a]	1	0.8

[a]Violent crimes.

Hypothesis 2: The Association Between Antenatal Depression and Antisocial Behavior Is Explained by Prior Environmental Adversity

Univariate analyses showed that, in these urban communities, depression in pregnancy was not significantly predicted by social class, cultural background, the mother's education, or the mother's IQ (see Table 3), but was more likely to occur if the mother was not legally married at the time of the child's birth, $\chi^2(1) = 3.74$, $p < .05$; if she had been a teenager at the time of the infant's birth, $\chi^2(1) = 3.77$, $p < .05$; and if she had been a teenager when she first entered into motherhood, at the birth of her first child, $\chi^2(1) = 6.04$, $p < .025$. However, the child's antisocial behavior was not predicted by any of these variables (see Table 4).

Violent adolescents did not differ significantly from other adolescents with respect to social class, maternal education, ethnicity, or maternal IQ, but were more likely to have been born to unmarried parents (cohabiting vs. married couples; none of the 8 adolescents born to single women was violent; see Table 5). When marital status was taken into account, depression in pregnancy remained a significant predictor of violent behavior, Wald statistic = 6.21, $p < .01$, OR = 3.87 (95% CI 1.34–11.20), and being born to unmarried parents was no longer significant.

Table 3

Associations Between Depression in Pregnancy and Other Potential Explanatory Variables

	Risk factor	Correlation
Antenatal mental health problems	Antenatal anxiety	.29***
Prior environmental adversity	Social class	.01
	Mother's education	.05
	Mother's IQ	−.14
	White British ethnicity	.02
	Parents unmarried	.18*
	One-parent family	.07
	Mother teenager this birth	.12*
	Mother teen at first birth	.21**
Child's re-exposure to depression	3 months postpartum	.33***
	12 months	.24**
	Years 1–4	.22*
	Years 4–11	.33***
Antenatal substance use	Cigarettes/day	.17†
	Units of alcohol/week	.02
Parents' antisocial behavior	Maternal CD symptoms	.20*
	Mother arrests	.10
	Father arrests	.12

Note. Kappa coefficients are reported for associations with nominal variables, point biserial coefficients for associations with continuous variables. IQ = intelligence quotient; CD = conduct disorder.

$^\dagger p < .10$. $^* p < .05$. $^{**} p < .01$. $^{***} p < .001$.

Hypothesis 3: The Effect of Depression in Pregnancy Is Explained by Later Exposure to the Mother's Depressive Illness

Mothers who were depressed in pregnancy were highly likely to become depressed again (see Table 3). Of the 38 women who were depressed in pregnancy, 16 (42.1%) were depressed postpartum, and 17 others (44.7%) became depressed at a later point in the child's life. All but 5 of the 38 women who had been depressed in pregnancy had at least one subsequent episode.

However, antisocial behavior was not associated with maternal depression at later points in the child's development, either in the months following childbirth, at 12 months of age, in early childhood, or in the primary school years (see Table 4). When depression at 3 months postpartum was included in the regression model, it was not significantly associated with antisocial behavior in adolescence, and depression during pregnancy remained a significant predictor of antisocial behavior. Nor was the effect of antenatal depression explained by the mother's subsequent depression, either at 12 months, in early childhood or the primary school years. Furthermore, antisocial adolescents did not differ significantly

Table 4

Exposure to Depression in Pregnancy and Other Risk Factors for Adolescents Who Had or Had Not Been Arrested or Diagnosed With Conduct Disorder (CD) at Age 11 and/or 16

	Risk factor	Antisocial (*N* = 40)	Not antisocial (*N* = 80)
Antenatal mental health problems	% Depressed*	45	25.0
	Anxiety, *M* (*SD*)	5.1 (3.2)	4.8 (3.2)
Prior environmental adversity	% Working class	87.5	88.8
	% Qualifications	75.0%	70.0%
	Mother's IQ, *M* (*SD*)	89.8 (12.9)	92.1 (14.2)
	% White British ethnicity	75.0	71.3
	% Parents unmarried	40.0	32.5
	% One-parent families	5.0	7.5
	% Teen this birth	7.5	6.3
	% Teen first birth	20.0	15.0
Re-exposure to depression	% Postpartum	22.5	21.3
	% 12 months	20.5	16.3
	% Years 1–4	35.0	31.3
	% Years 4–11	27.5	18.8
	Time points, median (range)	1 (0–4)	1 (0–4)
Substance use	Cigarettes/day, *M* (*SD*)	6.0 (9.0)	3.6 (6.0)
	Units/week, *M* (*SD*)	1.4 (3.6)	0.7 (1.2)
Parents' antisocial behavior	Maternal CD,** *M* (*SD*)	1.9 (2.3)	0.7 (1.3)
	% Mother arrests	5.3	6.3
	% Father arrests	42.5	36.3

Note. IQ = intelligence quotient.
*$p < .05$. **$p < .01$.

Table 5

Exposure to Depression in Pregnancy and Other Risk Factors for Adolescents Who Had or Had Not Been Arrested for Violent Crimes or Showed Aggressive Conduct Symptoms

	Risk factor	Violent (*N* = 18)	Not violent (*N* = 102)
Antenatal mental health problems	% Depressed**	61.1	26.5
	Anxiety, *M* (*SD*)	5.6 (4.1)	4.8 (3.0)
Prior environmental adversity	% Working class	83.3	89.2
	% Qualifications	61.1	73.5
	Mother's IQ, *M* (*SD*)	91.7 (15.6)	91.2 (13.5)
	% White British ethnicity	72.2	72.5
	% Parents unmarried*	55.6	31.4
	% One-parent families	0	7.8
	% Teen at this birth	11.1	5.9
	% Teen at first birth	27.8	14.7
Re-exposure to depression	% Postpartum*	33.3	19.6
	% 12 months	23.5	16.7
	% Years 1–4	38.9	31.4
	% Years 4–11*	38.9	18.6
	Time points,** median (range)	2 (0–4)	1 (0–4)
Substance use	Cigarettes/day, *M* (*SD*)	6.9 (7.7)	4.0 (7.1)
	Units/week, *M* (*SD*)	2.3 (5.0)	0.7 (1.3)
Parents' antisocial behavior	Maternal CD,* *M* (*SD*)	1.9 (2.3)	0.7 (1.3)
	% Mother arrests	12.5	5.9
	% Father arrests	44.4	37.3

Note. IQ = intelligence quotient; CD = conduct disorder.
*$p < .05$. **$p < .01$.

from other adolescents in the number of time periods during which they had been exposed to maternal depression, *Mdn* = 1 (range = 0–4) for the antisocial adolescents versus *Mdn* = 1 (range = 0–4) for the others, which was not a significant difference by the Mann–Whitney test, $z = -1.55$, *ns*.

Violent adolescents were somewhat more likely to have been exposed to the mother's depression between 4 and 11 years of age, $\chi^2(1) = 3.70$, $p < .054$ (see Table 5); the two groups did not differ with respect to exposure at other points in the child's lifetime. When exposure to maternal depression between 4 and 11 years was taken into account, depression in pregnancy remained a significant predictor of violence, Wald statistic = 5.37, $p < .02$, OR = 3.70 (95% CI 1.22–11.16), and the mother's depression in the middle childhood years no longer was significant. Adolescents who were violent had been exposed to maternal depression in a greater number of time periods, *Mdn* = 2 (range = 0–4) for violent youngsters versus *Mdn* = 1 (range = 0–4) for nonviolent ones, $z = -2.51$, $p < .01$ by the Mann–Whitney *U* test; however, this did not explain the effect of antenatal depression.

Hypothesis 4: The Effect of Depression in Pregnancy Is Explained by Mothers' Smoking and Drinking During Pregnancy

Women who were depressed in pregnancy smoked more cigarettes than women who were not depressed, *M* = 6.2 (*SD* = 7.1) versus *M* = 3.6 (*SD* = 7.2) for the depressed and nondepressed women, respectively, but the difference only approached statistical significance, $t(118) = -1.91$, $p < .06$. Depressed and nondepressed women did not differ in their reported use of alcohol during pregnancy, *M* = 1.0 (*SD* = 2.4) versus *M* = 0.9 (*SD* = 1.7), respectively.

Antenatal exposure to cigarette smoking and alcohol did not predict antisocial outcomes for the children. The mothers of antisocial teenagers had not smoked significantly more cigarettes per day and had not drunk significantly more units of alcohol per week, than other mothers in the sample (see Table 4). Similarly, the mothers of adolescents who had shown violent behavior had not been significantly more likely to smoke or drink alcohol in pregnancy (see Table 5). When fetal exposure to cigarette smoking and alcohol was taken into account, depression in pregnancy continued to predict antisocial behavior, Wald statistic = 3.83, $p < .05$, OR = 2.26 (95% CI 1.00–5.13), and violent behavior in particular, Wald statistic = 6.62, $p < .01$, OR = 4.07 (95% CI 1.40–11.84).

Hypothesis 5: The Association Between Depression in Pregnancy and Antisocial Behavior Is Explained by Associations With Biological Parents' Antisocial Behavior

To the extent that women's depression is often associated with a history of CD, and both disorders are influenced by genetic as well as environmental factors, the children of depressed mothers are at both genetic and environmental risk for antisocial outcomes. Furthermore, due to processes of assortative mating, children of depressed mothers who are also antisocial may be at genetic as well as environmental risk for antisocial behavior through processes of paternal as well as maternal transmission. We examined the evidence for each of these possibilities.

Associations between the mother's depression in pregnancy and her own past antisocial behavior. There was evidence for an association between depression in pregnancy and a juvenile history of CD. Women who were depressed in pregnancy reported significantly more juvenile conduct symptoms than did other women, *M* = 1.3 (*SD* = 1.9) versus *M* = 0.7 (*SD* = 1.3), respectively, which was significant by the Mann–Whitney *U* test, $z = 2.76$, $p < .01$. Women who had been depressed in pregnancy were also somewhat more likely than other women to have been arrested since the age of 18, 4 of 36 depressed women (11%) versus 3 of 81 nondepressed women (4%), but the difference was not significant.

Assortative mating. There was no clear evidence for assortative mating with respect to antisocial behavior. There was no significant association between the mothers' history of arrest and the biological fathers' history of arrest, kappa = .06, *ns*. Furthermore, when the biological father had a history of arrest, women were not more likely to report juvenile conduct symptoms, *M* = 1.0 (*SD* = 1.8) reported by women whose partners had been arrested and *M* = 0.8 (*SD* = 1.4) by women whose partners had not been arrested. Finally, there was no significant association between the mother's depression in pregnancy and the biological father's history of arrest; the father had been arrested in 18 of 38 families (47%) where the mother had been depressed in pregnancy and 28 of 82 families (34%) where she had not been depressed, which was not a significant difference.

Parents' antisocial behavior and adolescents' antisocial outcomes. Associations between parents' antisocial histories and the adolescents' antisocial behavior are presented in Table 4. Only 7 mothers had been arrested in adulthood and their history of arrest in adulthood was unrelated to adolescents' antisocial outcomes. However, the mothers' retrospective reports of antisocial behavior in their own youth did predict their children's antisocial outcomes. The mothers of antisocial adolescents reported $M = 1.4$ conduct symptoms ($SD = 1.9$) in their own teenage years, as opposed to $M = 0.7$ ($SD = 1.3$) reported by the other mothers, $z = -2.48$, $p < .01$ by the Mann–Whitney test. The mothers of adolescents who had shown violent behavior reported $M = 1.9$ conduct symptoms ($SD = 2.3$), as opposed to $M = 0.7$ ($SD = 1.3$) reported by the other mothers, $z = -2.79$, $p < .01$, by the Mann–Whitney test.

Forty-six of 120 biological fathers were known to have been arrested. Of those 46 antisocial men, 16 (34.8%) had offspring who were also antisocial. Of the 74 men who showed no evidence of antisocial behavior, 20 (27%) had biological children who were antisocial, which was not a significant difference.

Effects of depression in pregnancy, controlling for the mother's juvenile conduct symptoms. When the mothers' juvenile conduct symptoms were included in the regression model, neither depression in pregnancy nor the mother's conduct symptoms predicted the general measures of antisocial behavior; both were reduced to nonsignificant trends. However, depression in pregnancy remained a significant predictor of violent outcomes, Wald statistic = 5.71, $p < .02$, OR = 3.72 (95% CI 1.27–10.90), and the mothers' juvenile conduct symptoms independently predicted adolescents' violent behavior, Wald statistic = 4.57, $p < .03$. There was no evidence for an interaction between antenatal depression and maternal conduct symptoms.

When the biological fathers' history of arrest was taken into account, the association between depression in pregnancy and the adolescent's antisocial behavior remained significant, Wald statistic = 4.53, $p < .03$, OR = 2.41 (95% CI 1.07–5.40). Similarly, when the father's history of arrest was controlled for, the mother's depression in pregnancy continued to predict violence, Wald statistic = 7.37, $p < .01$, OR = 4.30 (95% CI 1.50–12.33).

Comparison of Violent and Nonviolent Antisocial Adolescents

The foregoing analyses showed that antenatal depression and the mother's history of conduct problems predicted violent outcomes despite the fact that nonviolent but antisocial youngsters were included in the comparison group for those analyses. Further analyses revealed that, compared to nonviolent antisocial youngsters and the adolescents who were not antisocial, the 18 adolescents who engaged in violent behavior had been exposed to higher levels of adversity in the social environment (Figure 1a). They were more likely to be exposed to maternal depression, particularly in pregnancy (Figure 1b). They also had been exposed to higher levels of a range of maternal risk factors (Figure 1c).

Planned contrasts compared the risk factors experienced by each group of antisocial adolescents (those who were violent, $N = 18$, and those who were not violent, $N = 22$) with the adolescents who were not antisocial. These analyses revealed that the violent youngsters differed significantly from adolescents who were not antisocial with respect to exposure to antenatal depression, $\chi^2(1) = 8.86$, $p < .01$, OR = 4.71 (95% CI 1.61–13.80); the number of time periods when the mother was depressed, $z = -2.87$, $p < .01$ by the Mann–Whitney test; and the number of maternal conduct symptoms, $z = -2.93$, $p < .01$ by the Mann–Whitney test. In contrast, the antisocial youngsters who were not violent did not differ significantly from adolescents who were not antisocial on any of these risk factors.

Sequential Analysis: A Pathway From Conduct Problems in One Generation to Violence in the Next

Because maternal conduct symptoms influenced violent outcomes in particular, in a final analysis, we asked how parental risk factors combined in sequence to predict violent behavior in adolescence. We sought evidence for an intergenerational pathway from the mother's juvenile conduct symptoms to her depression in pregnancy to the child's violent tendencies, and tested whether information about the biological father's antisocial behavior aided prediction of the adolescent's violence (see Figure 2). For the purpose of this final analysis, all variables were dichotomized, with women in the top quartile of conduct symptoms (≥ 1) compared with all others (see Table 6).

We first tested a first-order Markov chain model considering only relations between adjacent variables in this model: *Mother's conduct symptoms → Father's antisocial behavior → Mother's depression during pregnancy → Child's violence at 16 years.* This hypothesized order of events depicts the mother's history of antisocial behavior in her own adolescence, followed by assortative mating with an antisocial man, depression in pregnancy and

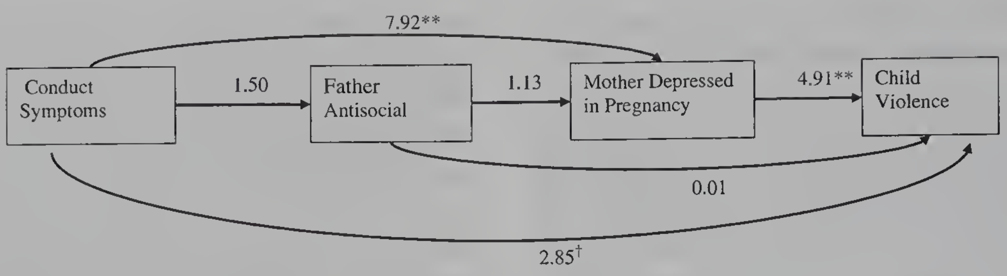


Figure 2. Second-order Markov chain analysis of sequences of risk events: the mother's conduct symptoms, the biological father's history of arrest, the mother's depression during pregnancy, and offspring violence at 16 years of age. In the picture, the numbers above the arrows indicate Wald tests of parameters for the relations between variables. The second-order model had a good fit to data: $\chi^2 = 3.09$, $p = .69$, likelihood ratio $L^2 = 4.14$, $p = .53$, $df = 5$, Bayesian information criterion = −19.79, Akaike information criterion = −5.86.
$^\dagger p < .10$. $^{**}p < .01$.

Table 6

Prediction of Violence at Age 16: Conditional Probabilities for Eight Different Combinations of Risk Events Leading to Violent Outcomes

Mother shows one or more conduct symptoms in adolescence	Biological father has been arrested	Mother becomes depressed during this pregnancy	Child is diagnosed with CD, with aggressive symptoms, and/or has committed violent crimes by 16 years	Conditional probability of violence, given preceding combination of events
Yes	Yes	Yes	Yes	.369
Yes	Yes	No	Yes	.146
Yes	No	Yes	Yes	.355
Yes	No	No	Yes	.138
No	Yes	Yes	Yes	.184
No	Yes	No	Yes	.062
No	No	Yes	Yes	.175
No	No	No	Yes	.058

Note. CD = conduct disorder.

violence in offspring. This first-order model did not have a good fit to the data observed ($\chi^2 = 14.53$, $p = .07$, likelihood ratio $L^2 = 15.29$, $p = .05$, $df = 8$, Bayesian information criterion [BIC] = −23.01, Akaike information criterion [AIC] = −0.71) and was therefore rejected.

We then tested the second-order Markov chain depicted in Figure 2. The results suggest that maternal conduct symptoms predicted depression in pregnancy, but only the mother's depression during pregnancy had a significant relation with the child's violence. Inspection of the conditional probabilities reveals additive effects of maternal conduct symptoms and antenatal depression (see Table 6). When both maternal conduct symptoms and antenatal depression were present, the conditional probability of the child being violent at 16 years, given both maternal risk factors, oscillated between 0.37, if the father had been arrested, and 0.36, if he had not been arrested. On the contrary, if mothers had not reported conduct symptoms, and had not become depressed in pregnancy, the chances of the child being violent at 16 were much lower, being 0.06 whether or not the father was antisocial.

Discussion

The foregoing analyses provide an illustration of ways in which to examine putative early experience effects while testing hypotheses about alternative explanations of those effects. Within a random sample of children born in two relatively disadvantaged urban communities in Britain, we tested for an association between maternal depression in pregnancy and antisocial outcomes in adolescence, and evaluated evidence for alternative hypotheses that might explain the association found. In general, mothers' depression in pregnancy placed their children at a twofold risk for antisocial outcomes, and at a fourfold risk for violent behavior, even when the alternative explanations were taken into account. The influence of depression in pregnancy was not explained by other environmental risk factors at the time of the pregnancy, or by exposure to the mother's illness at later time points, or by the mother's smoking cigarettes and drinking alcohol during the pregnancy. The mother's past history of conduct problems predicted her child's violent behavior but did not remove the effect of depression in pregnancy. Thus, after testing alternative hypotheses, the findings suggest that there was indeed an association between depression in pregnancy and violent antisocial behavior that was not explained by the competing hypotheses that were tested.

This evidence for an association does not, however, specify the causal mechanisms that might be at work. Our findings corroborate other studies linking pregnant women's mental health problems with disruptive behavior disorders and crime in offspring (Maki et al., 2003; O'Connor et al., 2002; Van den Bergh et al., 2006). Indeed, the findings are compatible with the proposition that there will be an increased prevalence of neurobiological deficits in antisocial youngsters, due to the effects of antenatal insults on fetal development (Raine, 2002; Raine et al., 2005; van Goozen et al., 2007). However, longitudinal prediction from antenatal depression to adolescent outcomes does not provide direct evidence for fetal programming. One major limitation of our study is the absence of information about fetal growth and neuroendocrine measures on mother and child.

Furthermore, the pattern of association among risk factors suggested that there might be psychological as well as biological sequelae to the mother's depression in pregnancy. The vast majority of children who were exposed to depression in pregnancy were re-exposed to maternal depression at a later point in their lives; during these subsequent episodes of mental illness, family life and opportunities for socialization may have been compromised. Although re-exposure to maternal depression did not explain the effects of depression in pregnancy, the subsequent episodes may well have conveyed incremental risk. The mother's own experience of depression during pregnancy may have colored her view of the fetus and her subsequent relationship with the child. Our data show that depression in pregnancy is a potent predictor of antisocial outcomes, particularly overtly violent behavior, but do not specify the underlying biological and psychological processes that are involved. It is of course possible that children whose self-regulatory abilities have failed to develop properly, due to early experiences before and after birth, may be particularly vulnerable to the later adversities that result from the mother's lifelong illness and antisocial tendencies.

We had hypothesized that the link between depression in pregnancy and the child's elevated risk for antisocial behavior would be linked to parents' own history of antisocial behavior. Our analyses confirmed that antenatal depression was associated with a history of conduct symptoms, but the mother's conduct symptoms did not explain the effect of antenatal depression. Children were at particular risk for violent behavior when their mothers had both a history of conduct problems and had been depressed in pregnancy. The father's history of arrest did not add predictive power. It would be unwise to conclude that paternal risk factors are unimportant, given that we did not have more detailed information about the father's own history of CD. Nevertheless, the main finding from these analyses was the importance of maternal CD and depressive illness for the development of adolescent antisocial behavior, and violence in particular.

It is possible that the importance of maternal disorder came to the fore in this sample because the sampling frame reduced variation in many of the common sociodemographic predictors of antisocial behavior, which also predict parents' psychopathology. Because the original aim of the study had been to examine perinatal mental illness in a sample where the prevalence of disorder was relatively high, these two urban communities had been chosen for the study (Sharp, 1993). Both the original assessments in the perinatal period and the follow-up assessments confirmed that symptoms of psychopathology and antisocial behavior were relatively common in both parents and offspring. A majority of children in the sample (65%) had been exposed to maternal depression at some point in their lives. In these urban communities, antisocial behavior was also relatively common. Over one third of the children had experienced a parent's arrest. One third of the adolescents themselves had manifested clinically and socially significant instances of antisocial behavior before the age of 16.

The socioeconomic characteristics of the SLCDS population may be seen as a limitation of the study. However, by drawing a random sample from a representative population that was at elevated risk for antisocial outcomes, we controlled for many of the broader environmental risk factors that are associated with antisocial behavior in more heterogeneous samples. Within the South London sample, broad measures of the social environment such as social class, maternal education, maternal IQ, and the family's cultural background did not predict antisocial outcomes. Even when we focused on the relatively small number of adolescents who engaged in overt violence, specific maternal characteristics (age at first birth, marital status, depression, and history of conduct problems), as opposed to the traditional social status measures of class, culture and education, were predictive.

Put another way, the study drew attention to the fact that the upper-middle-class children born in these urban communities were not necessarily protected from adverse outcomes. When compared to the working-class families (nearly 90% of the sample), the middle-class mothers were equally at risk for antenatal depression and their children were equally at risk for antisocial behavior. Our findings draw attention to the fact that not all working-class adolescents living in British cities are at risk for antisocial behavior. A focus on mothers' history of conduct problems and depression during pregnancy, as opposed to broader measures of the social environment, would hold promise for more targeted early interventions to prevent the development of serious antisocial behavior.

Another limitation of the study was its relatively small sample size; the design did not have the statistical power to detect small effects of antenatal influences that have been documented in larger surveys. For example, in this sample, neither smoking nor drinking in pregnancy exerted detectable effects on antisocial outcomes, and did not explain the association with antenatal depression.

It is noteworthy that the children who showed serious levels of violence in adolescence—having committed violent offences or shown violent conduct symptoms in the context of *DSM–IV* CD—were especially likely to have had a nonoptimal family environment. They were more likely than nonviolent children, even those with other antisocial symptoms, to have been exposed to recurring episodes of maternal illness and to have mothers with a history of conduct problems. The fact that antisocial adolescents who were not violent were not significantly different from non-antisocial adolescents in their exposure to most risk factors suggests that there may be a direct link between a child's early adverse experiences and his or her propensity for violence, as opposed to a disposition to show other forms of covert antisocial behavior (see Loeber et al., 1993). This in turn suggests that if there are fetal programming effects linked to the mother's depression in pregnancy, and leading to neurobiological impairment (Raine et al., 2005), they are more likely to influence overt aggression than other forms of rule breaking in childhood and adolescence. In the future, it might be helpful to test fetal programming hypotheses with respect to different categories of symptoms rather than a general diagnosis of CD or a summary measure of externalizing behaviors.

In sum, by testing a series of alternative hypotheses, we have identified two maternal variables that, in combination, led to high rates of antisocial behavior, and violence in particular. The importance of the additive risk of maternal conduct problems and antenatal depression was replicated across two different analytic approaches, logistic regression models and second-order Markov chain analysis. Although the relatively small sample size limited our ability to detect all risk factors that exert small effects on the development of antisocial behavior, the availability of detailed information on an initially random sample of mothers and children studied for 16 years provided a vivid picture of the challenges faced by the children of mothers with a history of conduct problems and lifelong depressive illness.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author

Angold, A., & Costello, E. J. (2000). The Child and Adolescent Psychiatric Assessment (CAPA). *Journal of the American Academy of Child and Adolescent Psychiatry*, *39*, 39–48.

Angold, A., Erkanli, A., Costello, E. J., & Rutter, M. (1996). Precision, reliability and accuracy in the dating of symptom onsets in child and adolescent psychopathology. *Journal of Child Psychology and Psychiatry*, *37*, 657–664.

Angold, A., Prendergast, M., Cox, A., Harrington, R., Simonoff, E., & Rutter, M. (1995). The Child and Adolescent Psychiatric Assessment (CAPA). *Psychological Medicine*, *25*, 739–753.

Brooker, B. H., & Cyr, J. J. (1986). Tables for clinicians to use to convert WAIS-R short forms. *Journal of Clinical Psychology*, *42*, 982–986.

Borders, A. E., Grobman, W. A., Amsden, L. B., & Holl, J. L. (2007). Chronic stress and low birthweight neonates in a low-income population of women. *Obstetrics and Gynecology*, *109*, 331–338.

Cairns, R. B. (1979). *Social development: The origins and plasticity of interchanges*. San Francisco: Freeman.

Campbell, S. B., & Cohn, J. F. (1997). The timing and chronicity of postpartum depression: Implications for infant development. In L. Murray & P. Cooper (Eds.), *Postpartum depression and child development* (pp. 165–197). New York: Guilford.

Carvalho De Moraes Barros, C. M., Guinsburg, R., Clovis De Araujo, P., Mitsuhiro, S., Chalem, E., & Laranjeira, R. R. (2006). Exposure to marijuana during pregnancy alters neurobehavior in the early neonatal period. *Journal of Pediatrics, 149*, 781–787.

Caspi, A., Taylor, A., Smart, M., Jackson, J., Tagani, S., & Moffitt, T. E. (2001). Can women provide reliable information about their children's fathers? Cross-informant agreement about men's lifetime antisocial behavior. *Journal of Child Psychology and Psychiatry, 42*, 915–920.

Cassidy, B., Zoccolillo, M., & Hughes, S. (1996). Psychopathology in adolescent mothers and its effect on mother–infant interactions: A pilot study. *Canadian Journal of Psychiatry, 41*, 379–384.

Dodge, K. A., Bates, J. E., & Pettit, G. S. (1990). Mechanisms in the cycle of violence. *Science, 250*, 1678–1683.

Galbaud du Fort, G., Boothroyd, L. J., Bland, R. C., Newman, S. C., & Kakuma, R. (2002). Spouse similarity for antisocial behavior in the general population. *Psychological Medicine, 32*, 1407–1416.

Goldberg, D., Cooper, B., Eastwood, M. T., Kedward, H. B., & Shepherd, M. (1970). A psychiatric interview for use in community surveys. *British Journal of Preventive Social Medicine, 24*, 18–23.

Goldthorpe, J. H., & Hope, K. (1974). *The social grading of occupations*. Oxford, UK: Oxford University Press.

Halligan, S. L., Murray, L., Martins, C., & Cooper, P. (2006). Maternal depression and psychiatric outcomes in adolescent offspring: A 13-year longitudinal study. *Journal of Affective Disorders, 97*, 145–154.

Hannaford, P. C., Smith, B. H., & Eliot, A. M. (2006). Primary care epidemiology: Its scope and purpose. *Family Practice*, 23, 1–7.

Hay, D. F., Pawlby, S., Angold, A., Harold, G., & Sharp, D. (2003). Pathways to violence in the children of mothers who were depressed post partum. *Developmental Psychology, 39*, 1083–1094.

Hay, D. F., Pawlby, S., Sharp, D., Asten, P., Mills, A., & Kumar, E. (2001). Intellectual problems shown by 11-year-old children whose mothers had postnatal depression. *Journal of Child Psychology and Psychiatry, 42*, 871–889.

Jaffee, S. R., Caspi, A., Moffitt, T. E., Belsky, J., & Silva, P. (2001). Why are children born to adolescent mothers at risk for adverse outcomes in young adulthood? Results from a 20-year longitudinal study *Development and Psychopathology, 13*, 377–397.

Keenan, K., & Shaw, D. (1994). The development of aggression in toddlers: A study of low-income families. *Journal of Abnormal Child Psychology, 22*, 53–77.

Kim-Cohen, J., Moffitt, T. E., Taylor, A., Pawlby, S. J., & Caspi, A. (2005). Maternal depression and child antisocial behavior: Nature and nurture effects. *Archives of General Psychiatry, 62*, 173–181.

Krueger, R. F., Moffitt, T. E., Caspi, A., Bleske, A., & Silva, P. A. (1998). Assortative mating for antisocial behavior: Developmental and methodological implications. *Behavior Genetics, 28*, 173–185.

Kupersmidt, J. B., Griesler, P. C., DeRosier, M. E., Patterson, C. J., & Davis, P. W. (1995). Childhood aggression and peer relations in the context of family and neighborhood factors. *Child Development, 66*, 360–375.

Loeber, R., Wung, P., Keenan, K., Giroux, B., Stouthamer-Loeber, M., & van Kammen, W. B. (1993). Developmental pathways in disruptive child behavior. *Development and Psychopathology, 5*, 101–132.

Lundy, B. L., Jones, N. A., Field, T., Nearing, G., Davalos, M., Pietro, P. A., et al. (1999). Prenatal depression effects on neonates. *Infant Behavior and Development, 22*, 119–129.

Maki, P., Veijula, J., Rasanen, P., Jouleamaa, M., Valonen, P., Jokelainen, J., et al. (2003). Criminality in the offspring of antenatally depressed mothers: A 33-year follow-up of the Northern Finland 1966 Birth Cohort. *Journal of Affective Disorders, 74*, 273–278.

Mathews, T. J., Menaker, F., & MacDorman, M. F. (2003). Infant mortality statistics from 2001 period linked birth/infant death data set. *National Vital Statistics Report, 52*, No. 2.

Maughan, B., Taylor, C., Taylor, A., Butler, N., & Bynner, J. (2001). Pregnancy smoking and childhood conduct problems: A causal association? *Journal of Child Psychology and Psychiatry, 42*, 1021–1028.

Moffitt, T. E. (2005). The new look of behavioral genetics in developmental psychopathology: Gene–environment interplay in antisocial behaviors. *Psychological Bulletin, 131*, 533–554.

National Health Service. (2004/2005). *Maternity statistics*. London: Department of Health.

O'Connor, T. G., Heron, J., Golding, J., Beveridge, M., & Glover, V. (2002). Maternal antenatal anxiety and children's behavioral/emotional problems at 4 years. Report from the Avon Longitudinal Study of Parents and Children. *British Journal of Psychiatry, 180*, 502–508.

O'Hara, M. W. (1997). The nature of postpartum depressive disorders. In L. Murray & P. J. Cooper (Eds.), *Postpartum depression and child development* (pp. 3–31). New York: Guilford.

Office for National Statistics. (2006). Available at http://www.statistics.gov.uk

Pawlby, S., Hay, D. F., Sharp, D., Waters, C. S., & O'Keane, V. (2009). Antenatal depression predicts depression in adolescent offspring: Prospective longitudinal community-based study. *Journal of Affective Disorders, 113*, 236–241.

Rahman, A., Bunn, J., Lovel, H., & Creed, F. (2007). Association between antenatal depression and low birthweight in a developing country. *Acta Psychiatrica Scandinavica, 115*, 481–486.

Raine, A. (2002). Annotation: The role of prefrontal deficits, low autonomic arousal, and early health factors in the development of antisocial and aggressive behavior in children. *Journal of Child Psychology and Psychiatry, 43*, 417–434.

Raine, A., Moffitt, T. E., Caspi, A., Loeber, R., Stouthamer-Loeber, M., & Lynam, D. (2005). Neurocognitive impairments in boys on the life-course persistent antisocial path. *Journal of Abnormal Psychology, 114*, 38–49.

Romano, E., Zoccolillo, M., & Paquette, D. (2006). Histories of child maltreatment and psychiatric disorder in pregnant adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry*, *45*, 329–336.

Rutter, M. (1972). *Maternal deprivation reassessed*. Harmondsworth, UK: Penguin.

Rutter, M., Giller, M., & Hagell, A. (1998). *Antisocial behavior by young people*. Cambridge, UK: University of Cambridge Press.

Sharp, D. (1993) *Childbirth-related emotional disorders in primary care: A longitudinal prospective study*. Unpublished Ph.D. thesis, University of London.

Snaith, R. P., Bridge, G., & Hamilton, M. (1971). The Leeds Scales for the Self-Assessment of Anxiety and Depression. *British Journal of Psychiatry*, *128*, 156–165.

Solomon, R. L., & Lessac, M. S. (1968). A control group design for experimental studies of developmental processes. *Psychological Bulletin*, *70*, 145–160.

Spitzer, R. L., Endicott, J., & Robins, E. (1978). *The Schedule for Affective Disorders and Schizophrenia–Lifetime version (SADS-L)* (3rd ed.). New York: Biometric Research.

Talge, N. M., Neal, C., Glover, V., & the Early Stress, Translational Research Prevention Science Network: Fetal and Neonatal Experience on Child Adolescence Mental Health. (2007). Antenatal maternal stress and long-term effects on child neurodevelopment: How and why? *Journal of Child Psychology and Psychiatry*, *48*, 245–261.

Van den Bergh, B. R. H., & Marcoen, A. (2004). High antenatal maternal anxiety is related to ADHD symptoms, externalizing problems, and anxiety in 8- and 9-year-olds. *Child Development*, *75*, 1085–1097.

Van den Bergh, B. R. H., Mennes, M., Oosterlaan, J., Stevens, V., Stiers, P., Marcoen, A., et al. (2005). High antenatal maternal anxiety is related to impulsivity during performance tasks in 14- to 15 year-olds. *Neuroscience and Biobehavioral Reviews*, *29*, 259–269.

van Goozen, S. H. M., Fairchild, G., Snoek, H., & Harold, G. T. (2007). The evidence for a neurobiological model of childhood antisocial behavior. *Psychological Bulletin*, *133*, 149–182.

Van den Bergh, B. R. H., Mennes, M., Stevens, V., Van der Meere, J., Borger, N., Stie, P., et al. (2006). ADHD deficit as measured in adolescent boys with a continuous performance task is related to antenatal maternal anxiety. *Pediatric Research*, *59*, 78–82.

Van den Bergh, B. R. H., Mulder, E. J. H., Mennes, M., & Glover, V. (2005). Antenatal maternal anxiety and stress and the neurobehavioral development of the fetus and the child: Links and possible mechanisms. A review. *Neuroscience and Biobehavioral Reviews*, *29*, 237–258.

Vermunt, J. K. (1997). *LEM 1.0: A general program for the analysis of categorical data*. Tilburg, Netherlands: Tilburg University.

Wechsler, D. (1981). *The Wechsler Adult Intelligence Scale Revised (WAIS-R)*. New York: Psychological Corporation.

Zoccolillo, M., Meyers, J., & Assiter, S. (1997). Conduct disorder, substance dependence, and adolescent motherhood. *American Journal of Orthopsychiatry*, *67*, 152–157.

Appendix: Three Cases Studies That Illustrate Some Individual Pathways to Violent Behavior in Adolescence

Nina: The Daughter of a Mother With Conduct Symptoms but No Depression in Pregnancy

Nina's mother, who was from a working class background, had a juvenile history of truancy and rule breaking. Neither she nor Nina's father had ever been arrested. At age 11, Nina had been oppositional and defiant toward the interviewer and had refused to do part of the interview. In secondary school, she spent little time at school and did not take any of the national GCSE examinations at age 16. By 16 years she had left school. She was arrested for fighting.

Victor: The Son of Antisocial Parents, With Maternal Depression in Pregnancy

Victor was the son of highly educated, upper-middle-class parents who had histories of juvenile conduct symptoms and arrest. Victor's mother, who was diagnosed with depression in pregnancy, had a juvenile history of truancy, breaking rules, arrest, running away, lying, alcohol and drug abuse, stealing, vandalism, and early sex. She had been arrested as a young woman for possession and use of drugs. Victor's father had been arrested for assault as a young man. Victor had a high IQ, had attended a private school, and had done well on the national GCSE examinations at age 16. He was diagnosed with conduct disorder, with aggressive symptoms (bullying, fighting, and use of a weapon), but had never been arrested.

Beatrice: The Daughter of an Antisocial Mother Who Became Depressed in Pregnancy

Beatrice's mother Renee had a juvenile history of truancy, rule breaking, running away, lying, and stealing. She was interviewed in the second trimester of pregnancy and diagnosed as being depressed. She regretted being pregnant and was worried that she would not be a good mother, despite being a mother of two other children. By age 16, Beatrice had been arrested 15 times and tried in court on five occasions on drug charges, for common assault, and for attempted robbery. At the time of the interview, Beatrice was worried about her pending court appearance, when she would either face being electronically tagged or imprisonment. (In the event, she was electronically tagged.) She told the interviewer, ''I don't care about anything, nothing at all. I was being evil, spiteful, getting in a lot of trouble from the police and stuff. I just didn't care.'' Beatrice had also met diagnostic criteria for an anxiety disorder at 11 and depression at 16.

Note. Names and identifying details have been changed to protect confidentiality.

Child Development, January/February 2010, Volume 81, Number 1, Pages 166–182

Exploring the Relation Between Prenatal and Neonatal Complications and Later Autistic-Like Features in a Representative Community Sample of Twins

Angelica Ronald, Francesca Happé, Katharina Dworzynski, Patrick Bolton, and Robert Plomin
Birkbeck College, King's College London

Prenatal and neonatal events were reported by parents of 13,690 eighteen-month-old twins enrolled in the Twins Early Development Study, a representative community sample born in England and Wales. At ages 7–8, parents and teachers completed questionnaires on social and nonsocial autistic-like features and parents completed the Childhood Asperger Syndrome Test. Correlations between prenatal and neonatal events and autistic-like features were weak, both in the whole sample (r = .00–.07) and at the 5% quantitative extreme (phenotypic group correlations = .01–.11), after controlling for socioeconomic status and cognitive ability. Neonatal problems showed modest heritability (13%–14%) and significant shared and nonshared environmental influences (55%–59% and 28%–31%, respectively). Differences in identical twins' neonatal problems correlated weakly with their difference scores on autistic-like features (r = .01–.06).

Obstetric factors and pre-, peri-, and neonatal problems have been found to be associated with autism and autistic spectrum disorders (ASD; Bolton et al., 1997; Brimacombe, Ming, & Lamendola, 2007; Burd, Severud, Kerbeshian, & Klug, 1999; Cederlund & Gillberg, 2004; Eaton, Mortensen, Thomsen, & Frydenberg, 2001; Gillberg & Cederlund, 2005; Glasson et al., 2004; Hultman, Sparen, & Cnattingius, 2002; Larsson et al., 2005; Lord, Mulloy, Wendelboe, & Schopler, 1991; Maimburg & Vaeth, 2006; Stein, Weizman, Ring, & Barak, 2006; for a selective review, see Kolevzon, Gross, & Reichenberg, 2007) although a number of studies have found little or no association (Cryan, Byrne, O'Donovan, & O'Callaghan, 1996; Deb, Prasad, Seth, & Eagles, 1997; Levy, Zoltak, & Saelens, 1988; Lord et al., 1991; Mason-Brothers et al., 1990; Piven et al., 1993). These findings of associations between obstetric and neonatal problems and autism alone do not necessarily denote causality and do not speak of the nature of the underlying risk factors. If prenatal and neonatal problems are causal risk factors for autism, it is not clear from existing data if these early complications are an environmental risk factor for autism, are a consequence of environmental insults occurring earlier in gestation, or if they are themselves genetically influenced and may be a result of existing genetic abnormalities in the fetus (Bolton et al., 1997; Gillberg, Gillberg, & Steffenburg, 1990; Goodman, 1990; Hultman & Sparen, 2004).

Twin studies have shown that autism is caused by both genetic influences and nonshared environmental influences (Bailey et al., 1995; Folstein & Rutter, 1977; Steffenburg et al., 1989). Nonshared environmental influences are environmental influences that make children growing up in the same family different and can include epigenetic processes, gene expression, de novo mutations, illnesses, intra- and extrauterine environment, and measurement error (Plomin, DeFries, McClearn, & McGuffin, 2008). In behavior genetics, many "environmental" variables have been shown to be under genetic influence, the so-called nature of nurture (Plomin & Bergeman, 1991). Using a classic twin design, it is possible to partition variance of a variable, such as birth weight or differential parenting,

---

We are indebted to participants of the Twins Early Development Study (TEDS) for making the study possible. The TEDS twin study was supported by a program grant MRC grant G0500079 from the UK Medical Research Council. P.B. was supported by the UK NIHR Biomedical Research Centre for Mental Health at the Institute of Psychiatry, Kings College London and The South London and Maudsley NHS Foundation Trust. A.R. was funded by an Autism Speaks fellowship.

Correspondence concerning this article should be addressed to Angelica Ronald, Centre for Brain and Cognitive Development, School of Psychology, Birkbeck College, London WC1E 7XH, UK. Electronic mail may be sent to a.ronald@bbk.ac.uk.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8101-0011

into genetic and environmental components to assess the extent to which it is explained by genetic and environmental influences.

In the original twin study of autism (Folstein & Rutter, 1977), biological hazards associated with the birth process, such as neonatal convulsions, did not explain the concordance rates for autism. However, in 12 of the 17 pairs discordant for autism, it was always the twin who had developed autism that had experienced a biological hazard (using a wide definition of biological hazards that included many features, e.g., multiple congenital abnormalities, delayed breathing after birth, a low birth weight compared to the unaffected cotwin, etc.). A similar pattern of findings was reported in the two later twin studies (Bailey et al., 1995; Steffenburg et al., 1989). In the fraternal or dizygotic (DZ) discordant twin pairs, the fact that the twin with autism showed more biological hazards associated with the birth process could be explained by environmental factors that differed between the twins or by genetic differences, as DZ twins share on average half their DNA (Plomin et al., 2008). However in the monozygotic (MZ) twins who were discordant for autism and for biological hazards at birth, these differences within a twin pair have to be explained by nonshared environmental processes (as defined above) because MZ twins share nearly all of their DNA code. This is why the discordant MZ twin design, which controls for most genetic influences, has proved a valuable tool for investigating whether associations are driven by nonshared environmental processes (e.g., Caspi et al., 2004; Newcombe, Milne, Caspi, Poulton, & Moffitt, 2007; Pike, Reiss, Hetherington, & Plomin, 1996).

As well as studying children with diagnosed autism, it is of interest to explore whether risk factors are present in siblings of children with autism. Several studies have shown that pregnancy and birth complications are associated not only in children with ASD but also in the siblings of children with ASD who show the broader autism phenotype (Bolton et al., 1997; Zwaigenbaum et al., 2002). This suggests that pregnancy and birth complications might be associated with subclinical autistic-like traits as well as with the diagnosed disorder.

A relatively recent approach has been to study dimensional measures of autistic behaviors throughout the range of typical variation and at the quantitative extreme using a general population sample (e.g., Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001; Constantino et al., 2003; Ronald, Happé, & Plomin, 2005). As well as allowing relation to be explored quantitatively rather than just within categorical clinical groups, which have restricted variance and their own sample biases (Caron & Rutter, 1991), another advantage of studying autistic-like features dimensionally in a general population sample is that it allows the different types of symptoms that make up the autism phenotype—such as social difficulties, communication problems, and nonsocial features such as repetitive behaviors—to be studied separately. Because the autism diagnosis requires children to show clinical-level social, communication, and nonsocial symptoms altogether, these different autistic features are by definition all shown at high levels and in combination with each other in samples of children with autism. A population sample offers the opportunity to study variation in autistic-like traits independently of each other (Happé, Ronald, & Plomin, 2006; Ronald, Happé, Bolton, et al., 2006).

Nevertheless, one study has explored the association between severity of scores on individual Autism Diagnostic Interview–Revised (ADI–R) and Autism Diagnostic Observation Schedule (ADOS) symptom domains and obstetric complications in a sample of children diagnosed with ASD (Wallace, Anderson, & Dubrow, 2008). This study did not find a relation between presence or absence of any obstetric variables and severity on the ADI–R social domain or the ADOS social/communication composite. However the authors did find that hypertension, pre-eclampsia, and generalized edema were all associated with the communication domain, and that hypertension, albuminuria, and generalized edema were associated with the repetitive behaviors domain. Although, as the authors note, it is difficult to measure the full range of severity within a sample of children with diagnosed ASD, and the ADI–R and ADOS are not designed to characterize severity of autistic symptoms, this initial study nevertheless highlights the possibility that obstetric and neonatal problems may show different degrees of association with the different symptoms that characterize autism.

The present study set out to investigate whether parent-rated prenatal and neonatal complications are associated with autistic-like features assessed dimensionally throughout the range of typical variation and at the quantitative extreme, and to explore the association between prenatal and neonatal complications and each type of autistic-like feature (social, communication, and nonsocial) separately, in a community sample. Second, by employing a classical twin design, we explored the extent to which neonatal problems are influenced by genetic and shared and nonshared environmental factors in the general population. Third, by utilizing an MZ differences design, we were able to control for most genetic effects on both neonatal problems and autistic-like features in order to explore whether the association between neonatal factors and autistic-like features in the general population was partly explained by nonshared environmental influences. As prenatal and neonatal problems and autistic-like features could be associated with lower socioeconomic status (SES) and lower cognitive ability in the general population (Newcombe et al., 2007; Hoekstra, Bartels, Verweij, & Boomsma, 2007), these two factors were regressed out of the scales to take account of their possible confounding effect in the analyses.

Method

Participants

The Twins Early Development Study (TEDS) is a UK-based community sample of twins (Oliver & Plomin, 2007; Trouton, Spinath, & Plomin, 2002). The Office for National Statistics (ONS) contacted the families of all live twin births between January 1994 and December 1996 that were recorded and traceable in England and Wales. ONS sent each family a pack including a letter (from ONS), an information leaflet (from TEDS) and a returnable postcard on which parents were asked to give consent for TEDS to contact them. ONS passed on to TEDS the names and addresses of the families that returned the postcard with consent: This was 65.1% of the families that ONS had attempted to contact and this initial sample consisted of 16,810 families in TEDS.

The first contact study was (in all but a few cases) the first point of contact between TEDS and the 16,810 families in the ONS sample. TEDS sent the first contact booklet to all except roughly 500 of the families who either withdrew from TEDS or provided ambiguous responses, had unclear contact details or replied too late for the first contact booklet.

The prenatal and neonatal problems data were collected in this first contact booklet, which was completed by the parent of the twins. The booklets were sent out when the twins were around 1½ years old (*M* = 1.87 years, *SD* = 1.27 years). 13,690 families (81.4% of those contacted) provided valid data at first contact.

Comparing the families that provided data on prenatal and neonatal problems at first contact to the UK population at that time, based on data from the Living in Britain report (ONS, 1997), 92% versus 93% were White, 49% versus 48% of the children were male, and 14% of mothers had one or more A-levels (UK advanced educational qualification) as their highest qualification compared to 9% of women in the United Kingdom.

Families were contacted again when the twins were 7 years old (*M* = 7.06, *SD* = 0.25): of the 14,581 families contacted, 1,679 families were no longer active, but of the active families, 8,163 (63% of active families at age 7; 59% of the families who provided first contact data) provided data at age 7. Parents were also asked for permission to contact the twins' teachers. Of the 7,524 teachers that were contacted, 6,156 (82%) returned data when the twins were age 7 (*M* = 7.20, *SD* = 0.28). The families were contacted approximately 1 year later when the twins were age 8 (*M* = 7.88, *SD* = 0.53); 13,940 families were contacted, 12,317 of whom were active and data were returned from 6,771 families (55% of active families at age 8; 49% of the families who provided first contact data).

The TEDS sample that provided data at ages 7 and 8 was reasonably representative of the UK population. For example, comparing the 7-year TEDS sample to data from the Living in Britain report (ONS, 1997), 94% versus 93% were White, 48% versus 48% were male, and 14% versus 9% of mothers had one or more A-level (UK advanced educational qualification) as their highest qualification. Four percent of children in the TEDS 7-year sample had a statement of special educational needs versus 3% of children in England (Department for Education and Skills, 2002). The same results were found for the sample with 8-year data. Finally, although clinical ASD diagnoses have not yet been completed, the approximate rate of likely ASD cases in this sample of 7,000 families based on the Diagnostic and Wellbeing Assessment (DAWBA) telephone interview instrument (see below) was 154 of 14,000 children (1.10%), which is comparable with the rate of 1.16% reported for all ASDs in a UK epidemiological study (Baird et al., 2006).

In terms of attrition, comparing those families participating at age 7 versus those families who only completed the first contact booklet and did not participate in TEDS subsequent to the first contact, 94% versus 88% were White, 14% versus 14% of mothers had A-levels as their highest educational qualification, and 45% versus 40% of mothers worked. Thus, the samples who provided first contact data and 7- and 8-year data were reasonably representative of the population and of the entire TEDS sample.

In 75% of cases, zygosity was diagnosed using DNA; for others, a parent-rated instrument was used (Price et al., 2000). Exclusions were made for specific medical syndromes such as Down syndrome, Fragile X, and chromosomal anomalies; unavailable zygosity data; no signed consent; missing birth order details; missing first contact data; and extreme perinatal outliers (gestation of < 27 weeks). The final samples after exclusions were 7,402 families with parent data at age 7, 5,982 families with teacher data at age 7, and 5,796 families with parent data at age 8.

Measures

Prenatal and neonatal problems. Parents completed a demographic questionnaire that asked about prenatal issues (12 questions about pregnancy, six about the labor) and neonatal variables (10 questions). The questionnaire included both continuous variables (degree of smoking and alcohol intake during pregnancy, the length of the labor, the length of gestation, the weight and length of the twins, time spent in hospital and in special care, and number of days breastfed) and categorical variables such as whether the mother had amniocentesis during pregnancy (see the Appendix for a full list of items). The questionnaire was based on the prenatal and neonatal reports employed in longitudinal studies at that time, such as the Avon Longitudinal Study of Pregnancy and Childhood and the MacArthur Longitudinal Twin Study, as well as findings from perinatal complications in psychopathology (Parnas et al., 1982) and based on the details that are acquired through Personal Child Health Records from hospitals.

Autistic-like features and potential ASD cases. Parents completed the Childhood Asperger Syndrome Test (CAST; Scott, Baron-Cohen, Bolton, & Brayne, 2002; Williams, Allison, et al., 2006; Williams, Scott, et al., 2005) when the twins were age 8. The CAST is a questionnaire designed to assess autistic-like behaviors for parents to complete in nonclinical settings. The CAST shows good internal consistency in this sample (alpha = .73) and it has been shown that parent reports on the CAST show acceptable test–retest reliability both in a high-scoring sample across a 2-month interval ($r = .67$; Allison et al., 2007) and across a 1-year interval in a community sample ($r = .63$; Ronald, Happé, & Plomin, 2008). Each of the 31 questions is answered in a *yes* (1) or *no* (0) format. The CAST items can be divided into three subscales, based on *DSM–IV* criteria (American Psychiatric Association, 1994; Ronald, Happé, Bolton, et al., 2006).

Parents and teachers also rated the twins' autistic-like features on *DSM–IV*-based social and nonsocial scales (Ronald et al., 2005) when the twins were age 7. The questionnaire was devised to assess social and nonsocial behaviors relevant to autism. The questionnaire was designed to include items that were relevant for assessing the types of social and nonsocial behaviors notable in autism but that also would be seen in the general population. The majority of items were derived from the *DSM–IV* autism criteria (see Ronald et al., 2005). The social scale had 10 items and the nonsocial scale had six items. Each item was rated as *not true* (0), *somewhat true* (1) or *certainly true* (2).

The *DSM–IV*-based social and nonsocial scales showed moderate internal consistency with Cronbach's alphas of .72 and .51 for teacher social and nonsocial scales, and .56 and .53 for parent social and nonsocial scales, respectively. Test–retest data were obtained for these scales from parents of a small independent sample of children aged 9–10, with a 1-week interval. The social scale showed good test–retest reliability ($r = .80$, $N = 35$); the nonsocial scale showed modest test–retest reliability ($r = .50$, $N = 35$). Test–retest data were not available for teacher report. Another index of reliability is the MZ twin correlation because reliability creates an upper limit for MZ twin correlations. MZ twin correlations for both the social and nonsocial scales were moderate to high for both parent (.63–.74) and teacher ratings (.60–.77).

Correlations between teachers and parents were moderate for the social scale ($r = .30$, $p < .001$) and low for the nonsocial scale ($r = .09$; $p < .001$). Similar levels of low to modest agreement have been reported in community samples between teacher and parent ratings of behavior problems (Achenbach, McConaughy, & Howell, 1987), have been observed in samples of children with ASD (Szatmari, Archer, Fisman, & Streiner, 1994) and autistic-like features (Ronald et al., 2008). For this reason the parent and teacher data were analyzed separately.

Because TEDS is a community sample representative of the general population, it is likely to include children with ASDs. Suspected ASD cases have been identified (though not yet clinically assessed) in TEDS via the CAST questionnaire, which acts as a screening tool, or by parents alerting TEDS directly that their child had an ASD, or if parents ticked an ASD box in a questionnaire. Parents of the suspected cases were interviewed by telephone using the DAWBA (Goodman, Ford, Richards, Gatward, & Meltzer, 2000). Potential ASD cases ($N = 154$; 52% Autism, 9% Aspergers, 39% Other Pervasive Developmental Disorder) were identified from these telephone interviews; the twin concordances are reported elsewhere (Bolton et al., 2009). These children were included in the analyses but analyses were also repeated without these DAWBA-ascertained ASD cases.

Other measures. SES was derived from five measures collected in the first contact booklet: highest qualification of mother, highest qualification of father, social class group of father (ranging from unskilled occupation to professional), social class group of mother, and age of mother at the birth of her first child. A single socioeconomic status (SES) composite was created by standardizing these five variables and taking an average score (Asbury, Dunn, & Plomin, 2006). General cognitive ability was assessed at age 7 using Similarities, Vocabulary, and Picture Completion subtests from the Wechsler Intelligence Scale for Children (Wechsler, 1992) and the Conceptual Grouping subtest from the McCarthy Scales of Children's Abilities (McCarthy, 1972; see also Davis & Plomin, 2008).

Analyses. All scales were created by summing items equally and converting to proportions of the total possible score given the number of items completed (which had to be more than half). If there were data for fewer than half the items, the total score was coded as missing.

The categorical items about prenatal and neonatal problems (see the Appendix) were all coded 1 for *yes* and 0 for *no* and were added together into a Prenatal and Neonatal Problems (PNP) scale. This included 17 child-specific and non-child-specific categorical prenatal and neonatal problem items such as "Did you experience high blood pressure during your pregnancy?" (non-child-specific item) and "Did the twin have special care when born (e.g., incubators, etc.)?" (child-specific item). The continuous measures (number of cigarettes smoked during pregnancy, units of alcohol drank during pregnancy, gestation, time in labor, birth length, birth weight, time in hospital, days in special care, days breastfed) were analyzed individually in relation to autistic-like features.

For the twin analyses, a separate Neonatal Problems (NP) scale was created which only included the child-specific categorical variables (see the Appendix) and was created, separately for Twin 1 and Twin 2. High scores on the NP scale indicated a greater number of postnatal child-specific neonatal problems for an individual twin.

Phenotypic analyses. Phenotypic correlations and phenotypic group correlations (Oliver, Dale, & Plomin, 2004) were used to assess the strength of the relation between prenatal and neonatal problems and autistic-like features throughout the general population and at the extreme, respectively. Phenotypic correlations were obtained using STATA 10.0 (StataCorp. 2007). The robust cluster option was used, which corrects standard errors to take account of the nonindependence of data when multiple family members are included (Huber, 1967). Phenotypic group correlations were calculated by dividing the proband's mean standardized score on the unselected variable by the proband's mean standardized score on the selected variable, where probands are defined as all children scoring above a 95% cutoff. Phenotypic group correlations are thus bidirectional: the phenotypic group correlation representing the extreme prenatal and neonatal problems proband group's scores on autistic-like features is different from the phenotypic group correlation representing the extreme autistic-like features group's scores on prenatal and neonatal problem scales. If the regression of one variable on another is linear across the full sample, then the phenotypic group correlation is expected to be similar in magnitude to the correlation for the entire sample (Plomin, Price, Eley, Dale, & Stevenson, 2002) .

Extreme features were defined as scores in the top 95% range of the distribution. This cutoff was chosen as a compromise between the need for a sample size large enough to provide adequate power versus the desire to study extreme scoring groups. General cognitive ability and SES were regressed out of the PNP and NP scales to control for their effects. Similarly, gestation was regressed out of the birth weight variable.

Univariate twin model of neonatal problems. The twin design is based on comparing the within-pair similarity of MZ and DZ twins on a measure or trait of interest (Plomin et al., 2008). The following analyses involved partitioning the variance in neonatal problems in this sample into genetic and environmental influences. Twin analyses can only be carried out on variables that are specific to each child; therefore, neonatal problems, which differ for each child, were studied using a twin model but variables that are shared between twins, such as gestation length, could not be used in the twin model fitting. In accordance with standard quantitative genetics procedure, the neonatal problem scale was corrected for sex prior to analyses. The first step in such analyses was to explore twin similarity coefficients (intraclass correlations; Shrout & Fleiss, 1979). Next, an ACE univariate model was employed, so called because it partitions the total variance into three components: additive genetic (*A*), shared environment (*C*), and nonshared environment (*E*).

Heritability refers to the proportion of variation of a trait in a population explained by genetic influences. *Environmental influences* in the twin design refers to all variance that is not explained by genetic influences (i.e., it has a broader definition than the lay meaning of the environment), and is split into two types, shared and nonshared. Shared environment refers to experiences that make children growing up in the same family similar; nonshared environment refers to environmental influences that make children growing up in the same family different and can include epigenetic processes, gene expression, de novo mutations, illnesses, intra- and extrauterine environment, and measurement error (Plomin et al., 2008).

A univariate twin model partitions variance in a measure or trait into genetic and environmental influences. It does this by observing the degree of covariance shown between twin pairs, separately for MZ and DZ twins. The model assumes that MZ twins share all their genetic influences and all their shared environment, and that DZ twins share on average half their genes and all their shared environment. No measured genes or measured environments are included in the model.

Model fitting was carried out using raw data and the *Mx* program to provide parameter estimates for *A*, *C*, and *E* as well as confidence intervals (Neale, Boker, Xie, & Maes, 2003). The likelihood ratio $\Delta\chi^2$ test (LRT), the Akaike's information criterion, and sample-size-adjusted Bayesian information criterion were used to compare the fit of different models. The LRT uses the log likelihood statistic −2LL with associated *p* values, and selects the most parsimonious model, that is, the model with the best fit given the number of degrees of freedom.

The type of univariate model that was employed here included tests for three types of sex differences and is called a univariate sex limitation model. The model can test, through a series of submodels, whether the same genetic and shared environmental influences affect a trait for males and females (qualitative sex differences), whether genetic and environmental influences have a different degree of magnitude in their influence for males and females (quantitative sex differences), and whether there are variance differences for males and females.

In the univariate sex limitation model the genetic coefficient (r_G) represents genetic relatedness between twins in a pair (Neale & Cardon, 1992). In same-sex MZ groups, r_G is fixed at 1 because MZ twins are assumed to share all their genetic makeup; in same-sex DZ groups, r_G is fixed at .5 because DZ twins share, on average, half their segregating genes. In the DZOS group, r_G is allowed to vary below .5 (the value ascribed to it in DZ same-sex groups) in order to test for qualitative sex differences. The shared environmental coefficient (r_C) is fixed at 1 for MZ and DZ same-sex groups because it is assumed that same-sex twins growing up in the same family share all their shared environmental influences. However, r_C is allowed to vary below this value in the DZOS group. By allowing r_G and r_C to be lower in the DZOS group than in DZSS groups, this tests the extent of genetic and shared environmental overlap between males and females, called qualitative sex differences. Quantitative sex differences can be assessed by allowing *A*, *C*, and *E* parameters to vary for male and female groups. The third sex difference involving variance differences between males and females is assessed by allowing female variance components to be equal to a scalar multiple of the male variance components.

In the full univariate sex limitation model, qualitative and quantitative sex differences are tested. The raw data ACE full model has 15 free parameters: *A* (male), *A* (female), *C* (male), *C* (female), *E* (male), *E* (female), r_G or r_C, male and female variance, and male and female means for MZ, DZ same-sex, and DZ opposite-sex groups. The common effects submodel tests for quantitative sex differences but it sets r_G and r_C at their fixed values (as in the same-sex DZ groups). The scalar submodel tests for sex differences in variance only: it allows the female variance components to be equal to a scalar multiple of the male variance components. Finally, the null submodel fixes all parameters (except for the means) to be equal for males and females.

MZ differences analyses. Difference scores were calculated between MZ twins by subtracting Twin 1's score from Twin 2's score (see Pike et al., 1996). Pearson correlations were then derived for the relation between the difference scores for the prenatal and neonatal variables, and the difference scores for autistic-like features.

Results

A multiple regression of all the prenatal and neonatal problem individual items (listed in the Appendix) as independent variables and autistic-like features as dependent variables using a forced entry method showed that the prenatal and neonatal problem items together explained 5% of the variance in the total CAST score at age 8, and 4%, 4%, and 3% of the variance in the social, communication, and nonsocial subscales at age 8, respectively. At age 7, prenatal and neonatal problems explained 4% of variance in parent-rated social autistic-like features and 2% of variance in parent-rated nonsocial features, and 2% of variance in both the social and nonsocial teacher scales.

Table 1 shows the phenotypic correlations between continuous prenatal and neonatal variables and autistic-like features at ages 7 and 8, which range from ±.00 to ±.10 in magnitude. The PNP scale and the NP scale both showed very low but significant positive correlations with both parent and teacher ratings of both social and nonsocial autistic-like features at age 7 ($r = .03$–$.07$), with the exception of the teacher nonsocial scales that had nonsignificant correlations. Both the PNP scale and the NP scale also showed positive and significant correlations with the parent-rated CAST scales at age 8 ($r = .04$–$.07$), this time with the exception of social autistic-like features, which did not have significant correlations with either PNP or NP scales.

Of the other continuous prenatal and neonatal problem variables in Table 1, length of breastfeeding and number of cigarettes smoked during pregnancy showed the most number of significant associations: Days breastfed correlated negatively with most, though not all, of the autistic-like feature scales. Number of cigarettes during pregnancy correlated positively and significantly with all autistic-like features ($r = .03$–$.10$) except for teacher-rated nonsocial behaviors at age 7. For most of the other continuous prenatal and neonatal variables, correlations were mostly nonsignificant but showed a consistent trend: Gestation correlated for the most part negatively with autistic-like features, birth weight correlated negatively with autistic-like features (with the exception of the CAST social scale), time in hospital correlated positively with autistic-like features (significant for most scales at age 7 but not at age 8), and days in special care correlated positively but weakly with autistic-like features. For alcohol drunk during pregnancy, birth length, and time in labor, correlations varied between positive and negative and were for the most part nonsignificant. Correlations were highly similar after removing the small proportion of children with DAWBA-rated ASDs from the sample.

Table 1

Correlations Between Prenatal and Neonatal Variables and Autistic-Like Features After Correcting for Social Class and General Cognitive Ability

	7 years				8 years			
	Social (parent)	Social (teacher)	Nonsocial (parent)	Nonsocial (teacher)	Total CAST	Social	Comm	Nonsocial
Prenatal and Neonatal Problems (PNP) scale	.07**	.06**	.04**	.00	.07**	.02	.07**	.06**
Neonatal Problems (NP) scale	.05**	.06**	.03*	.02	.05**	.02	.05**	.04**
Cigarettes smoked	.10**	.03*	.06**	.00	.09**	.06**	.07**	.05**
Alcohol drank	.00	.00	−.03*	−.02	−.01	−.02	.00	−.01
Gestation	−.03**	−.03*	−.02	−.03	−.01	.02	−.03	−.01
Time in labor	.02	.00	.03*	−.01	.01	−.01	.01	.02
Birth weight[a]	−.01	.00	−.02	−.01	−.01	.03**	−.03*	−.02
Birth length	−.02	.00	−.02	−.02	.03	.05*	.01	.00
Days in special care	.03	.04	.03	.01	.01	.00	.02	.00
Time in hospital	.04**	.06**	.03*	.03	.02	.00	.03*	.01
Days breastfed	−.03*	.03	−.04*	.00	−.05**	−.03*	−.03	−.05**

Note. See the Appendix for descriptions of PNP and NP measures. General cognitive ability was assessed in the children at age 7; see the Method section for details of the measure. These correlations take into account relatedness between twin pairs. CAST = Childhood Asperger Syndrome Test; Social = social autistic-like features; Comm = communication autistic-like features; Nonsocial = nonsocial behaviors.

[a] Corrected for gestational age.

*$p < .05$. **$p < .01$.

Table 2 shows the phenotypic group correlations between prenatal and neonatal variables and autistic-like features selected at the quantitative extreme (using a 5% cutoff to define probands), which ranged from ±.00 to ±.11. Overall the strength of the relation between prenatal and neonatal variables and autistic-like features at the extreme was of a magnitude similar to the strength of the relation between prenatal and neonatal variables and autistic-like features throughout the whole sample. For both the PNP scale and the NP scale, there tended to be a stronger relation with autistic-like features at age 8 when selecting the autistic-like feature probands and looking at their PNP or NP scores, than for selecting the children with extreme prenatal and neonatal problems and investigating their scores for autistic-like features.

For the variables in Tables 1 and 2, there were some significant mean sex differences between male and female twins (opposite-sex twins were omitted from these comparisons), with males scoring higher on the PNP scale, $F(1,\ 5225) = 12.34$, $p < .001$, and the NP scale, $F(1,\ 5234) = 26.52$, $p < .001$, and males showing a greater birth length, $F(1,\ 2397) = 6.91$, $p = .01$, and greater birth weight (corrected for gestation), $F(1,\ 5067) = 83.79$, $p < .001$. Time in hospital, days breastfed, days in special care, time in labor, number of cigarettes smoked, and units of alcohol drunk during pregnancy did not differ significantly for males and females. Phenotypic correlations between autistic-like features and the prenatal and neonatal variables did not differ significantly for males and females.

Twin correlations for the NP scale are shown in Table 3. MZ correlations were higher than DZ correlations (although confidence intervals overlapped), suggesting a small amount of genetic influence on the NP scale. MZ correlations were significantly lower than unity, reflecting modest nonshared environmental influences on this scale. DZ correlations were considerably greater than half the MZ correlations, suggesting significant shared environmental influences on this scale. MZ and DZ same-sex correlations differed slightly for males and females, suggesting some small quantitative

Table 3
Twin Correlations for Neonatal Problems Scale

		N
MZM	.72 (.68–.75)	772
DZM	.63 (.59–.67)	735
MZF	.68 (.64–.71)	868
DZF	.60 (.56–.65)	775
DZOS	.53 (.49–.57)	1,445

Note. MZM = monozygotic male; MZF = monozygotic female; DZM = dizygotic male; DZF = dizygotic female; DZOS = opposite-sex DZ twins.

Table 2
Phenotypic Group Correlations Between Prenatal and Neonatal Variables and Autistic-Like Features, Using a 95% Cutoff, After Regressing Out SES and g

	7 years				8 years			
	Social (parent)	Social (teacher)	Nonsocial (parent)	Nonsocial (teacher)	Total CAST	Social	Comm	Nonsocial
Prenatal and Neonatal Problems (PNP) scale	.06/.06	.05/.03	.06/.05	.01/.01	.11/.02	.06/−.02	.08/.04	.11/.04
Neonatal Problems (NP) scale	.04/.02	.04/.01	.02/.08	.04/.03	.08/.01	.04/.01	.07/−.01	.05/.03
Cigarettes smoked	.11/.11	.02/.03	.02/.05	.01/−.02	.11/.08	.09/.06	.07/.06	.07/.05
Alcohol drank	.01/.00	.01/−.01	−.04/−.02	−.02/−.01	−.03/−.05	−.03/−.04	−.02/−.03	−.04/−.04
Gestation	−.02/−.05	−.02/−.06	.02/−.06	−.02/−.04	−.03/.03	.02/.07	−.04/−.01	−.03/−.01
Time in labor	.01/.00	.01/.02	.00/.02	.01/.01	.01/−.03	.02/−.05	.00/−.02	.01/−.01
Birth weight[a]	−.02/−.02	−.01/−.06	−.03/−.05	.01/−.03	−.04/.01	.02/.04	−.04/.00	−.03/−.02
Days in special care	−.02/.03	−.01/.07	−.03/.06	−.02/.05	−.01/−.05	−.05/−.08	−.01/−.04	−.01/.02
Time in hospital	.03/.01	.04/.04	.02/.02	.03/.02	.04/−.02	.00/−.03	.05/−.01	.04/−.01
Days breastfed	−.05/.16	.03/.10	−.03/.08	−.01/−.22	−.02/.25	−.03/.11	−.03/.28	−.03/.15

Note. See the Appendix for descriptions of PNP and NP measures. Autistic-like feature as selected variable/Prenatal or neonatal measure as selected variable. CAST = Childhood Asperger Syndrome Test; Social = social autistic-like features; Comm = communication autistic-like features; Nonsocial = nonsocial behaviors.
[a] Corrected for gestational age.

sex differences. Finally, the DZOS correlation was significantly lower than the DZ same-sex correlations, implicating that there might be qualitative sex differences on neonatal problems. Twin correlations were highly similar after removing the children with DAWBA-based ASDs from the sample.

The results of the univariate sex limitation model are presented in Table 4. The best fitting model was an ACE full sex limitation model, which meant that both qualitative and quantitative sex differences were significant. As suggested from the twin correlations, the NP scale showed low but significant heritability (13% for males, 14% for females), significant shared environment (59% for males, 55% for females), and modest nonshared environmental influence (28% for males, 31% for females). Parameter estimates could not be equated for males and females. r_G could not be fixed at 0.5 without a significant decrease in fit and was estimated at .00 (.00–.20); r_C was estimated at .85 (.79–.91), demonstrating that there were significant qualitative sex differences on the scale.

Table 5 presents the correlations between MZ pair difference scores for prenatal and neonatal variables with MZ pair difference scores for autistic-like features, which ranged from ±.00 to ±.10. For both the PNP scale and the NP scale, there were weak positive correlations between MZ pair difference scores for prenatal and neonatal problems with MZ pair difference scores for autistic-like features (.01–.06). However, not all these correlations were significant: Of the 7-year scales, only the MZ differences scores in parent-rated social autistic-like features showed significant correlations with the MZ differences scores in the PNP scale (.05) and the NP scale (.05), and of the 8-year scales only the MZ differences scores in the total CAST scale and the communication autistic-like features subscale showed significant correlations with the MZ differences scores in the PNP scale and the NP scale (.05–.06). MZ differences scores in birth weight and time in hospital both showed significant correlations with the MZ differences scores in social but not nonsocial autistic-like features at age 7; similar results were suggested with the 8-year data but the correlations were not significant. MZ differences correlations between days breastfed and autistic-like features as well as between days in special care and autistic-like features were for the most part nonsignificant.

Discussion

This is the first study to explore the relation between autistic-like features, measured dimensionally in the general population and at the quantitative extreme, and parent-reported prenatal and neonatal events. Our hypothesis was that prenatal and neonatal complications, which show an elevated rate in children with ASD, would also show an association with autistic-like features assessed dimensionally in the general population. The analyses also explored whether prenatal and neonatal

Table 4
Results From Univariate Sex Limitation Model for Neonatal Problems Scale

Model	−2LL	*Par*	*df*	χ^2 difference (LRT)	AIC	Sample-size-adjusted BIC	*p*		a^2	c^2	e^2
Saturated	24,612.75	25	9,536								
ACE full model	24,627.71	13	9,548	14.96	−9.04	−13,046.49	.244	Male:	.13 (.05–.20)	.59 (.52–.66)	.28 (.25–.31)
								Female:	.14 (.06–.21)	.55 (.48–.61)	.31 (.29–.35)
ACE common effects model	24,637.60	12	9,549	24.85	−1.15	−13,044.20	.024				
ACE scalar model	24,690.05	10	9,551	77.30	47.30	−13,023.29	< .001				
ACE null model	24,713.10	9	9,552	100.36	68.36	−13,014.42	< .001				

Note. Estimated genetic correlation between opposite-sex dizygotic twin pairs (indicating qualitative sex differences) = .00 (.00–.20); estimated shared environment correlation between opposite-sex dizygotic twin pairs (indicating qualitative sex differences) = .85 (.79–.91). A/a^2 = additive genetic influences; C/c^2 = shared environmental influences; E/e^2 = nonshared environmental influences; −2LL = log likelihood fit statistics; LRT = likelihood ratio test; AIC = Akaike's information criterion; BIC = Bayesian information criterion.

Table 5

Correlations Between MZ Pair Differences in Autistic-Like Features and MZ Pair Differences in Neonatal Variables

	7 years (*N* [teacher] = 1,969–2,041; *N* [parent] = 2,524–2,619)				8 years (*N* = 2,005–2,080)			
	Social (parent)	Social (teacher)	Nonsocial (parent)	Nonsocial (teacher)	Total CAST	Social	Comm	Nonsocial
PNP	.05*	.03	.01	.03	.05*	.03	.06*	.02
NP	.05*	.03	.01	.03	.06**	.04	.06**	.02
Birth weight[a]	−.07**	−.10**	.03	.00	−.04	−.02	−.04	−.02
Time in hospital	.04*	.05*	.00	.01	.03	.03	.01	.02
Days breastfed	.00	−.01	.02	−.01	.03	.01	.02	.03
Days in special care	.05*	.01	.01	.03	.02	.05*	.02	−.03

Note. See the Appendix for descriptions of PNP and NP measures. Removed twin pairs with ±> 3.5 *SD* difference in scores. CAST = Childhood Asperger Syndrome Test; Social = social autistic-like features; Comm = communication autistic-like features; Nonsocial = nonsocial behaviors; PNP = Prenatal and Neonatal Problems scale; NP = Neonatal Problems scale.

[a] Corrected for gestational age.

*$p < .05$. **$p < .01$.

complications are more strongly associated with some specific autistic-like features (e.g., social impairments) more than others (e.g., nonsocial autistic-like behaviors). A twin design was employed to explore to what degree an association between prenatal and neonatal events and autistic-like features may be explained by shared genetic influences or by environmental processes.

Strengths and Limitations

We begin with a discussion of the strengths and limitations of the study. First, in terms of limitations, we relied on parent report for our measure of prenatal and neonatal problems collected on average 1½ years after the twins' birth. It has been shown that maternal report shows very good agreement with medical records for information on the child's birth weight, admission to special care baby unit, the method of delivery, maternal smoking during pregnancy, high blood pressure during pregnancy, and breastfeeding, but not for alcohol use during pregnancy and length of labor (Buka, Goldstein, Spartos, & Tsuang, 2004; Githens, Glass, Sloan, & Entman, 1993; Li, Scanlon, & Serdula, 2005; Rice et al., 2007; Sanderson et al., 1998; Seidman, Slater, Ever-Hadani, & Gale, 1987; Tate, Dezateux, Cole, & Davidson, 2005; Tilley et al., 1985; Walton et al., 2000). Maternal characteristics such as educational qualifications have been shown to not influence the level of agreement (Rice et al., 2007). Nevertheless, it would be useful in future research on this issue to obtain hospital records concerning prenatal and neonatal history. Information about placentation was not available in the present study but has proved important in, for example, the study of birth problems in schizophrenia (Davis, Phelps, & Bracha, 1995). Children were not examined for visible dismorphic features that may be associated with prenatal or neonatal history.

Concerning the measurement of autistic-like features, quantitative measures of autistic-like features are still in development—there is no gold standard as there is with ASD diagnoses—and although they may be more appropriate for tapping individual differences in variation than clinical tools, their validity is not as clear-cut. The sensitivity of the CAST as a screening tool for ASD with a designated cut-point of 15 has been shown, in a sample of children clinically assessed using the ADI–R and ADOS, to be 100%, the specificity 97%, but the positive predictive value 50% (Williams et al., 2005). Unfortunately data on the validity the *DSM–IV* scales used at age 7 in our sample, by comparing them to ADOS and ADI–R scores in a clinical sample, are not yet available. Although it has been shown that measures of autistic-like traits are to some extent independent of other forms of psychopathology (e.g., Constantino, Hudziak, & Todd, 2003; Hoekstra et al., 2007), and, as described above, the CAST shows excellent specificity, it is nevertheless possible that parents of children with other disorders endorsed the autistic-like feature items. Therefore, when considering how these findings have relevance to clinical ASD, it is important that they are considered in light of the properties of the measures that were used.

Other considerations include the fact that sampling autistic-like features more proximally in time, such as in early childhood, to the prenatal and neonatal events might have reduced noise and led to higher correlations. This would have made possible a bivariate twin model-fitting analysis to examine directly the degree of shared genetic and environmental effects on neonatal problems and autistic-like features. Possible confounding variables in the intervening years between the pregnancy and birth and the assessment of autistic-like features should be considered, and we did not have information on birth order within the family, so we were not able to take that into account.

Examining a twin sample clearly brings significant advantages; however, it also involves important constraints. Twin analyses can only be carried out on variables that are specific to each child; those that are shared between them, such as gestation length, cannot be used in model fitting. As both twins share most intrauterine environmental exposures, the range of child-specific problems that could be examined using the twin design was narrowed to neonatal experiences in these data. In general, twins also show a much greater incidence of prenatal and neonatal problems than singletons, which means that they are an enriched sample for studying such problems, yet it might also mean that prenatal and neonatal problems in twins do not have as much predictive power as the same level of complications in singletons. Therefore, it is important to give careful consideration to how these findings from a twin sample are generalizable to nontwin samples. The conclusions reached from the twin analyses on neonatal problems might not apply to prenatal or obstetric factors. The twin design assumes that MZ twins are 100% similar genetically, but studies have shown that there are some small differences in DNA in identical twins, for example, differences in copy number variations (Bruder et al., 2008). Finally, like all research designs, the twin design is subject to its assumptions and limitations and it is optimal to try to triangulate on these issues with other research designs such as family and adoption studies (Plomin et al., 2008).

Set against the limitations, the strengths of this study include the fact that data on prenatal and neonatal events were collected in the child's 2nd year, considerably before the data on autistic-like features and therefore can be considered somewhat independent of parents' later concerns about their children's social development (although it is possible that some of these problems may have begun to emerge at the time the prenatal and neonatal data were collected). The longitudinal nature of the study allowed us to test for associations between prenatal and neonatal problems and later autistic-like features that endured across early and middle childhood. Although it remains possible that parent bias explains the relation between parent-rated prenatal and neonatal events and parent-rated autistic-like features, by also including some teacher ratings of autistic-like features we were able to partly circumvent the problem of common methods variance and rater bias. Having two different measures of autistic-like features and two different raters in the present study allowed us to attempt to internally replicate our findings so that they were not specific to a particular measure or rater. However, this strength in the research design left us with some inconsistencies in the results, that is, some significant effects that were not consistent across both measures and raters, and meant that the analyses were subject to multiple testing.

Finally, having a large population sample allowed us to analyze quantitative extreme scoring groups as well as the whole sample. We found that the strength of the association between prenatal and neonatal problems and autistic-like features was consistent across the whole sample and at the quantitative extreme. It should be reiterated that these findings stem from data in a community sample, not a clinical ASD sample. The analyses on "extreme" groups, which were based on a 95% cutoff, will include children who are not as extreme as a diagnosed sample. A second advantage of the large sample was that it provided us with power to detect small effects. Importantly, a review of studies of measured nonshared environmental variables concluded that they tend to account for only small portions of variance in behavior (Turkheimer & Waldron, 2000), which highlights the need for large samples to have power to detect these nonshared environmental effects.

Interpretation of Results

Prenatal and neonatal problems accounted for a small amount (2%–5%) of the variance in autistic-like features. The relation between prenatal and neonatal problems and autistic-like features was weak across the typical range and at the extreme and, when tested using two different measures of autistic-like features and using two different raters, did not provide consistently significant results. Further research is needed to confirm whether the significant results are correct.

Prenatal and neonatal problems did not show a consistently stronger relation with some types of autistic-like features more than others. Correlations did not differ significantly, which may in part be due to their small absolute magnitude. Nevertheless, the data demonstrate the utility of exploring associations with the different types of autistic-like features separately in a general population sample, especially given previous evidence suggesting that different autistic-like features may have different underlying genetic and environmental influences (Happé et al., 2006; Ronald, Happé, Bolton, et al., 2006; Ronald, Happé, Price, Baron-Cohen, & Plomin, 2006).

It is important to note that prenatal and neonatal complications are not exclusively associated with autism and autistic-like features: They have been shown to be associated with intelligence quotient (Newcombe et al., 2007) and many other forms of psychopathology (Eaton et al., 2001). This was reflected in our data in the phenotypic group correlations that, when selecting on children with prenatal and neonatal problems and investigating their autistic-like features scores, showed a much weaker relation than when selecting children with extreme autistic-like features and investigating their prenatal and neonatal experiences. Although we were able to control for the possible confounding effects of cognitive ability and SES in the analyses, it is possible that other correlated medical comorbidities could explain the covariance between autistic-like features and prenatal and neonatal events.

The aim of employing a twin design in this study was to go beyond looking for an association and to explore the extent to which the relation between neonatal problems and autistic-like features might be due to common genetic and environmental influences. The twin model of neonatal problems showed that neonatal problems were explained by low but significant heritability and both shared and nonshared environmental influences. Previous twin studies on these same data have shown that autistic-like features are caused by genetic influences and nonshared environmental influences (Ronald et al., 2005; Ronald, Happé, Bolton, et al., 2006). Previously, there have been suggestions in the autism literature that the association with neonatal problems in autism reflects preexisting genetic abnormality (Bolton et al., 1997; Goodman, 1990). The present study provides data to support this assertion insomuch as we found that genetic influences play a small but significant role in causing variation in neonatal problems in the general population. At a purely speculative level, if there are shared genetic influences between autistic-like features and neonatal problems, this would suggest that some of the genes involved in risk for developing autistic-like features begin to be expressed at the time of birth and affect the developing child in such a way as to increase the risk of problems at birth as well as their later behavior. Such a finding would aid our understanding of the underlying neurobiological pathways involved in the development of autistic-like features. However, in our data neonatal problems were very weakly (and not always significantly) phenotypically associated with autistic-like features and these genetic pathways are only speculative at this stage.

One puzzle about the findings is that twin studies of autism and autistic-like traits suggest, for the most part, that shared environment plays little or no role in their etiology (e.g., Bailey et al., 1995; Folstein & Rutter, 1977; Ronald, Happé, Bolton, et al., 2006; Steffenburg et al., 1989), yet some prenatal factors, such as maternal smoking (which are necessarily shared between both twins in a pair), showed a significant association. One explanation could be that the effect size of these shared environmental risk factors, such as maternal smoking, is small. Another explanation is that prenatal risk factors interact with genetic liability via gene–environment interaction or gene–environment correlation but are not a risk factor on their own.

The key evidence for "nonshared" environmental influences underlying the weak association between neonatal problems and autistic-like features in the general population can be taken from the data on discordant MZ twins. Within this general population sample, differences within MZ pairs in neonatal problems showed a weak (although not consistent) association with differences within MZ pairs in later autistic-like features. The strength of this approach is that these MZ differences correlations are not likely to be caused by shared genetic influences because the comparisons are made within MZ twin pairs, who share most of their DNA. If the results were more consistent, these data would suggest that there are some common nonshared environmental effects causing both neonatal problems and autistic-like features. For example, a nongenetically determined fetal abnormality might predispose a neonate to a more hazardous postnatal course, reflected in both more neonatal problems and later autistic-like features. However, it should be noted that not all of the MZ differences correlations were significant and therefore further evidence is needed for this nonshared environmental association.

Therefore, these analyses explored the possibility that *both* genetic and nonshared environmental influences are involved in causing the association between prenatal and neonatal problems and later autistic-like features, as measured here. A further future step after establishing that both genetic influences and environment may play a role in causing neonatal problems is to investigate the possibility of gene–environment interactions as it is unlikely that these two processes operate in isolation.

In terms of sex differences, mean sex differences were observed in some of the prenatal and neonatal variables, most notably the PNP scale and the NP scale, on which males scored significantly higher than females. This is in line with previous reports showing that males have more birth defects and more adverse neonatal outcomes than females (e.g., Lary & Paulozzi, 2001; Stevenson et al., 2000). Furthermore, the lower female than male MZ correlations for the NP scale suggest the possible involvement of the X chromosome in influencing neonatal problems (Loat, Asbury, Galsworthy, Plomin, & Craig, 2004). Autistic-like features have been shown elsewhere to show significant mean sex differences, with males scoring higher than females (e.g., Baron-Cohen et al., 2001; Ronald, Happé, Bolton, et al., 2006). The univariate twin model of neonatal problems suggested there were some small qualitative and quantitative sex differences in the genetic and environmental causes of neonatal problems. Future research could explore further this link between the sex bias in prenatal and neonatal problems and the sex bias in autistic-like features and ASDs.

In sum, these results provide some support for a weak association between prenatal and neonatal events and later autistic-like features in the general population and at the quantitative extreme. Unfortunately not all associations were found to be significant. Previous reports have shown an association between prenatal and neonatal problems and diagnosed autism. On balance, further research is needed to confirm whether the association also stretches to subthreshold autistic-like features. Neonatal problems showed a small amount of genetic influence in the general population, suggesting that the association between neonatal problems and later autistic-like features could be partly due to shared genetic influences, although this could not be tested directly because of the low correlations. Within-pair differences in identical twins' neonatal experiences correlated weakly with later within-pair differences in autistic-like features, suggesting there may also be common nonshared environmental influences explaining a small amount of the covariance between neonatal problems and later autistic-like features. Again, conclusions are tempered by the inconsistencies in the results. Finally there was not consistent evidence that prenatal and neonatal problems show a stronger association with some autistic-like features more than others. As well as the need for replication, an interesting question for future research will be to explore how the different kinds of prenatal and neonatal experiences that were summed together here, and that are likely to have different etiological components, might contribute individually to risk for autistic-like features.

References

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101,* 213–232.

Allison, C., Williams, J., Scott, F., Stott, C., Bolton, P., Baron-Cohen, S., et al. (2007). The Childhood Asperger Syndrome Test (CAST): Test–retest reliability in a high scoring sample. *Autism, 11,* 173–185.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Asbury, K., Dunn, J. F., & Plomin, R. (2006). Birthweight-discordance and differences in early parenting relate to monozygotic twin differences in behaviour problems and academic achievement at age 7. *Developmental Science, 9,* F22–F31.

Baird, G., Simonoff, E., Pickles, A., Chandler, S., Loucas, T., Meldrum, D., et al. (2006). Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: The Special Needs and Autism Project (SNAP). *Lancet, 368*(9531), 210–215.

Bailey, A., Le Couteur, A., Gottesman, I., Bolton, P., Simonoff, E., Yuzda, E., et al. (1995). Autism as a strongly genetic disorder: Evidence from a British twin study. *Psychological Medicine, 25,* 63–77.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders, 31,* 5–17.

Bolton, P. F., Dworzynski, K., Curran, S., Ronald, A., Happé, F., Goodman, R., et al. (2009). *A population twin study of autistic spectrum disorders.* Manuscript in preparation.

Bolton, P. F., Murphy, M., Macdonald, H., Whitlock, B., Pickles, A., & Rutter, M. (1997). Obstetric complications in autism: Consequences or causes of the condition?

Journal of the American Academy of Child & Adolescent Psychiatry, 36, 272–281.

Brimacombe, M., Ming, X., & Lamendola, M. (2007). Prenatal and birth complications in autism. *Maternal and Child Health Journal, 11*, 73–79.

Bruder, C. E., Piotrowski, A., Gijsbers, A. A., Andersson, R., Erickson, S., de Stahl, T. D., et al. (2008). Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. *American Journal of Human Genetics, 82*, 763–771.

Buka, S. L., Goldstein, J. M., Spartos, E., & Tsuang, M. T. (2004). The retrospective measurement of prenatal and perinatal events: Accuracy of maternal recall. *Schizophrenia Research, 71*, 417–426.

Burd, L., Severud, R., Kerbeshian, J., & Klug, M. G. (1999). Prenatal and perinatal risk factors for autism. *Journal of Perinatal Medicine, 27*, 441–450.

Caron, C., & Rutter, M. (1991). Comorbidity in child psychopathology: Concepts, issues and research strategies. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 32*, 1063–1080.

Caspi, A., Moffitt, T. E., Morgan, J., Rutter, M., Taylor, A., Arseneault, L., et al. (2004). Maternal expressed emotion predicts children's antisocial behavior problems: Using monozygotic-twin differences to identify environmental effects on behavioral development. *Developmental Psychology, 40*, 149–161.

Cederlund, M., & Gillberg, C. (2004). One hundred males with Asperger syndrome: A clinical study of background and associated factors. *Developmental Medicine and Child Neurology, 46*, 652–660.

Constantino, J. N., Davis, S. A., Todd, R. D., Schindler, M. K., Gross, M. M., Brophy, S. L., et al. (2003). Validation of a brief quantitative measure of autistic traits: Comparison of the social responsiveness scale with the autism diagnostic interview-revised. *Journal of Autism and Developmental Disorders, 33*, 427–433.

Constantino, J. N., Hudziak, J. J., & Todd, R. D. (2003). Deficits in reciprocal social behavior in male twins: Evidence for a genetically independent domain of psychopathology. *Journal of the American Academy of Child & Adolescent Psychiatry, 42*, 458–467.

Cryan, E., Byrne, M., O'Donovan, A., & O'Callaghan, E. (1996). A case–control study of obstetric complications and later autistic disorder. *Journal of Autism and Developmental Disorders, 26*, 453–460.

Davis, J. O., Phelps, J. A., & Bracha, H. S. (1995). Prenatal development of monozygotic twins and concordance for schizophrenia. *Schizophrenia Bulletin, 21*, 357–366.

Davis, O. S., & Plomin, R. (2008). g in middle childhood: Moderate genetic and shared environmental influence using diverse measures of general cognitive ability at 7, 9 and 10 years in a large population sample of twins. *Intelligence, 36*, 68–80.

Deb, S., Prasad, K. B., Seth, H., & Eagles, J. M. (1997). A comparison of obstetric and neonatal complications between children with autistic disorder and their siblings. *Journal of Intellectual Disability Research, 41*(Part 1), 81–86.

Department for Education and Skills (2002). *Special educational needs in England: January, 2002 bulletin.*

Eaton, W. W., Mortensen, P. B., Thomsen, P. H., & Frydenberg, M. (2001). Obstetric complications and risk for severe psychopathology in childhood. *Journal of Autism and Developmental Disorders, 31*, 279–285.

Folstein, S., & Rutter, M. (1977). Infantile autism: A genetic study of 21 twin pairs. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 18*, 297–321.

Gillberg, C., & Cederlund, M. (2005). Asperger syndrome: Familial and pre- and perinatal factors. *Journal of Autism and Developmental Disorders, 35*, 159–166.

Gillberg, C., Gillberg, C., & Steffenburg, S. (1990). Reduced optimality in the pre-, peri-, and neonatal periods is not equivalent to severe peri- or neonatal risk: A rejoinder to Goodman's technical note. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 31*, 813–815.

Githens, P. B., Glass, C. A., Sloan, F. A., & Entman, S. S. (1993). Maternal recall and medical records: An examination of events during pregnancy, childbirth, and early infancy. *Birth, 20*, 136–141.

Glasson, E. J., Bower, C., Petterson, B., de Klerk, N., Chaney, G., & Hallmayer, J. F. (2004). Perinatal factors and the development of autism: A population study. *Archives of General Psychiatry, 61*, 618–627.

Goodman, R. (1990). Technical note: Are perinatal complications causes or consequences of autism? *Journal of Child Psychology & Psychiatry & Allied Disciplines, 31*, 809–812.

Goodman, R., Ford, T., Richards, H., Gatward, R., & Meltzer, H. (2000). The Development and Well-Being Assessment: Description and initial validation of an integrated assessment of child and adolescent psychopathology. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 41*, 645–655.

Happé, F., Ronald, A., & Plomin, R. (2006). Time to give up on a single explanation for autism. *Nature Neuroscience, 9*, 1218–1220.

Hoekstra, R. A., Happé, F., Baron-Cohen, S., & Ronald, A. (in press). The association between extreme autistic traits and intellectual disability: Insights from a general population twin study. *British Journal of Psychiatry.*

Hoekstra, R. A., Bartels, M., Verweij, C. J., & Boomsma, D. I. (2007). Heritability of autistic traits in the general population. *Archives of Pediatrics and Adolescent Medicine, 161*, 372–377.

Huber, P. J. (1967). *The behavior of maximum likelihood estimates under non-standard conditions.* Paper presented at the proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability, Berkeley, CA.

Hultman, C. M., & Sparen, P. (2004). Autism–prenatal insults or an epiphenomenon of a strongly genetic disorder? *Lancet, 364*(9433), 485–487.

Hultman, C. M., Sparen, P., & Cnattingius, S. (2002). Perinatal risk factors for infantile autism. *Epidemiology, 13*, 417–423.

Kolevzon, A., Gross, R., & Reichenberg, A. (2007). Prenatal and perinatal risk factors for autism: A review and integration of findings. *Archives of Pediatrics and Adolescent Medicine, 161*, 326–333.

Larsson, H. J., Eaton, W. W., Madsen, K. M., Vestergaard, M., Olesen, A. V., Agerbo, E., et al. (2005). Risk factors for autism: Perinatal factors, parental psychiatric history, and socioeconomic status. *American Journal of Epidemiology, 161*, 916–925.

Lary, J. M., & Paulozzi, L. J. (2001). Sex differences in the prevalence of human birth defects: A population-based study. *Teratology, 64*, 237–251.

Levy, S., Zoltak, B., & Saelens, T. (1988). A comparison of obstetrical records of autistic and nonautistic referrals for psychoeducational evaluations. *Journal of Autism & Developmental Disorders, 18*, 573–581.

Li, R., Scanlon, K. S., & Serdula, M. K. (2005). The validity and reliability of maternal recall of breastfeeding practice. *Nutrition Reviews, 63*, 103–110.

Loat, C. S., Asbury, K., Galsworthy, M. J., Plomin, R., & Craig, I. W. (2004). X inactivation as a source of behavioural differences in monozygotic female twins. *Twin Research, 7*, 54–61.

Lord, C., Mulloy, C., Wendelboe, M., & Schopler, E. (1991). Pre- and perinatal factors in high-functioning females and males with autism. *Journal of Autism and Developmental Disorders, 21*, 197–209.

Maimburg, R. D., & Vaeth, M. (2006). Perinatal risk factors and infantile autism. *Acta Psychiatrica Scandinavica, 114*, 257–264.

Mason-Brothers, A., Ritvo, E. R., Pingree, C., Petersen, P. B., Jenson, W. R., McMahon, W. M., et al. (1990). The UCLA-University of Utah epidemiologic survey of autism: Prenatal, perinatal, and postnatal factors. *Pediatrics, 86*, 514–519.

McCarthy, D. (1972). *McCarthy Scales of Children's Abilities*. New York: Psychological Corporation.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2003). *Mx statistical modeling* (6th ed.).

Neale, M. C., & Cardon, L. R. (1992). *Methodology for genetic studies of twins and families*. Dordrecht: Kluwer.

Newcombe, R., Milne, B. J., Caspi, A., Poulton, R., & Moffitt, T. E. (2007). Birthweight predicts IQ: Fact or artefact? *Twin Research and Human Genetics, 10*, 581–586.

Office for National Statistics. (1997). *Living in Britain: Results from the 1995 General Household Survey*. London: The Stationary Office.

Oliver, B. R., Dale, P. S., & Plomin, R. (2004). Verbal and non-verbal predictors of early language problems: An analysis of twins in early childhood back to infancy. *Journal of Child Language, 31*, 609–631.

Oliver, B. R., & Plomin, R. (2007). Twins' Early Development Study (TEDS): A multivariate, longitudinal genetic investigation of language, cognition and behavior problems from childhood through adolescence. *Twin Research Human Genetics, 10*, 96–105.

Parnas, J., Schulsinger, F., Teasdale, T. W., Schulsinger, H., Feldman, P. M., & Mednick, S. A. (1982). Perinatal complications and clinical outcome within the schizophrenia spectrum. *British Journal of Psychiatry, 140*, 416–420.

Pike, A., Reiss, D., Hetherington, E. M., & Plomin, R. (1996). Using MZ differences in the search for nonshared environmental effects. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 37*, 695–704.

Piven, J., Simon, J., Chase, G. A., Wzorek, M., Landa, R., Gayle, J., et al. (1993). The etiology of autism: Pre-, peri- and neonatal factors. *Journal of the American Academy of Child & Adolescent Psychiatry, 32*, 1256–1263.

Plomin, R., & Bergeman, C. S. (1991). The nature of nurture: Genetic influence on "environmental" measures. *Behavioral & Brain Sciences, 14*, 373–427.

Plomin, R., DeFries, J. C., McClearn, G. E., & McGuffin, P. (2008). *Behavioral genetics* (4th ed.). New York: Worth.

Plomin, R., Price, T. S., Eley, T. C., Dale, P. S., & Stevenson, J. (2002). Associations between behaviour problems and verbal and nonverbal cognitive abilities and disabilities in early childhood. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 43*, 619–633.

Price, T. S., Freeman, B., Craig, I., Petrill, S. A., Ebersole, L., & Plomin, R. (2000). Infant zygosity can be assigned by parental report questionnaire data. *Twin Research, 3*, 129–133.

Rice, F. L. A., Harold, G., van den Bree, M., Boivin, J., Hay, D. F., Owen, M. J., et al. (2007). Agreement between maternal report and antenatal records for a range of pre and peri-natal factors: The influence of maternal and child characteristics. *Early Human Development, 83*, 497–504.

Ronald, A., Happé, F., Bolton, P., Butcher, L. M., Price, T. S., Wheelwright, S., et al. (2006). Genetic heterogeneity between the three components of the autism spectrum: A twin study. *Journal of the American Academy of Child & Adolescent Psychiatry, 45*, 691–699.

Ronald, A., Happé, F., & Plomin, R. (2005). The genetic relationship between individual differences in social and nonsocial behaviours characteristic of autism. *Developmental Science, 8*, 444–458.

Ronald, A., Happé, F., & Plomin, R. (2008). A twin study investigating the genetic and environmental aetiologies of parent, teacher and child ratings of autistic-like traits and their overlap. *European Child and Adolescent Psychiatry, 17*, 473–483.

Ronald, A., Happé, F., Price, T. S., Baron-Cohen, S., & Plomin, R. (2006). Phenotypic and genetic overlap between autistic traits at the extremes of the general population. *Journal of the American Academy of Child & Adolescent Psychiatry, 45*, 1206–1214.

Sanderson, M., Williams, M. A., White, E., Daling, J. R., Holt, V. L., Malone, K. E., et al. (1998). Validity and reliability of subject and mother reporting of perinatal factors. *American Journal of Epidemiology, 147*, 136–140.

Scott, F. J., Baron-Cohen, S., Bolton, P., & Brayne, C. (2002). The CAST (Childhood Asperger Syndrome Test): Preliminary development of a UK screen for mainstream primary-school-age children. *Autism, 6*, 9–31.

Seidman, D. S., Slater, P. E., Ever-Hadani, P., & Gale, R. (1987). Accuracy of mothers' recall of birthweight and gestational age. *British Journal of Obstetrics and Gynaecology, 94*, 731–735.

Shrout, P. E., & Fleiss, J. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420–428.

StataCorp. (2007). *Stata (Release 10) [Computer software]*. College Station, TX: StataCorp.

Steffenburg, S., Gillberg, C., Hellgren, L., Andersson, L., Gillberg, I. C., Jakobsson, G., et al. (1989). A twin study of autism in Denmark, Finland, Iceland, Norway and Sweden. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 30*, 405–416.

Stein, D., Weizman, A., Ring, A., & Barak, Y. (2006). Obstetric complications in individuals diagnosed with autism and in healthy controls. *Comprehensive Psychiatry, 47*, 69–75.

Stevenson, D. K., Verter, J., Fanaroff, A. A., Oh, W., Ehrenkranz, R. A., Shankaran, S., et al. (2000). Sex differences in outcomes of very low birthweight infants: The newborn male disadvantage. *Archives of Disease in Childhood. Fetal and Neonatal Edition, 83*, F182–F185.

Szatmari, P., Archer, L., Fisman, S., & Streiner, D. L. (1994). Parent and teacher agreement in the assessment of pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 24*, 703–717.

Tate, A. R., Dezateux, C., Cole, T. J., & Davidson, L. (2005). Factors affecting a mother's recall of her baby's birth weight. *International Journal of Epidemiology, 34*, 688–695.

Tilley, B. C., Barnes, A. B., Bergstralh, E., Labarthe, D., Noller, K. L., Colton, T., et al. (1985). A comparison of pregnancy history recall and medical records. Implications for retrospective studies. *American Journal of Epidemiology, 121*, 269–281.

Trouton, A., Spinath, F. M., & Plomin, R. (2002). Twins Early Development Study (TEDS): A multivariate, longitudinal genetic investigation of language, cognition and behavior problems in childhood. *Twin Research, 5*, 444–448.

Turkheimer, E., & Waldron, M. (2000). Nonshared environment: A theoretical, methodological, and quantitative review. *Psychological Bulletin, 126*, 78–108.

Wallace, A. E., Anderson, G. M., & Dubrow, R. (2008). Obstetric and parental psychiatric variables as potential predictors of autism severity. *Journal of Autism and Developmental Disorders, 38*, 1542–1554.

Walton, K. A., Murray, L. J., Gallagher, A. M., Cran, G. W., Savage, M. J., & Boreham, C. (2000). Parental recall of birthweight: A good proxy for recorded birthweight? *European Journal of Epidemiology, 16*, 793–796.

Wechsler, D. (1992). *Wechsler Intelligence Scale for Children* (3rd ed.). London: Psychological Corporation.

Williams, J., Allison, C., Scott, F., Stott, C., Bolton, P., Baron-Cohen, S., et al. (2006). The Childhood Asperger Syndrome Test (CAST): Test–retest reliability. *Autism, 10*, 415–427.

Williams, J., Scott, F., Stott, C., Allison, C., Bolton, P., Baron-Cohen, S., et al. (2005). The CAST (Childhood Asperger Syndrome Test): Test accuracy. *Autism, 9*, 45–68.

Zwaigenbaum, L., Szatmari, P., Jones, M. B., Bryson, S. E., MacLean, J. E., Mahoney, W. J., et al. (2002). Pregnancy and birth complications in autism and liability to the broader autism phenotype. *Journal of the American Academy of Child & Adolescent Psychiatry, 41*, 572–579.

Appendix

Demographic Questionnaire to Assess Prenatal and Neonatal Variables

Non-child-specific items	Response format
Did you smoke cigarettes whilst pregnant?	Yes/No
How many cigarettes did you smoke each day, on average?	Number of cigarettes
Did you drink alcohol whilst pregnant?	Yes/No
How many units of alcohol did you drink, per week, on average? (1 unit = 1 glass of wine, OR, 1 measure of spirits, OR, ½ a pint of beer)	Number of units
Did you experience any severe stress during your pregnancy (for example, bereavement, serious illness in the family, major money problems)?[a]	Yes/No
During your pregnancy did you experience any of the following	
Amniocentesis[a]	Yes/No/Don't know
High blood pressure[a]	Yes/No/Don't know
Diabetes[a]	Yes/No/Don't know
Toxaemia/pre-eclampsia[a]	Yes/No/Don't know

Appendix Continued

Non-child-specific items	Response format
Vaginal bleeding[a]	Yes/No/Don't know
Rubella/German measles[c]	Yes/No/Don't know
Slow growth of baby/ies[a]	Yes/No/Don't know
Waters breaking more than 11 h before labor[a]	Yes/No/Don't know
How many weeks pregnant were you at the time of delivery?	Number of weeks
Was your labor induced?[a]	Yes/No/Don't know
How long was your labor?	Number of hours
Were there any complications during the birth of your twins?[a]	Yes/No/Don't know
Was the birth by Caesarean section?[a]	Yes expected/Yes emergency/No/Don't know

Child-specific items	
What were the weights of the twins at birth?	
1st born	In lb, oz or kg
2nd born	In lb, oz or kg
What were the lengths of the twins at birth?	
1st born	In cm or in.
2nd born	In cm or in.
Was there any concern about either twin at birth (e.g., did either twin have any breathing problems)?[ab]	
1st born	Yes/No/Don't know
2nd born	Yes/No/Don't know
Did either of the twins have any special care when born (e.g., incubators etc.)?[ab]	
1st born	Yes/No/Don't know
2nd born	Yes/No/Don't know
How long did they stay in special care?	
1st born	Number of days
2nd born	Number of days
How long did the twins stay in hospital after birth?	
1st born	Number of days
2nd born	Number of days
Has either of the twins had to go back to hospital as an outpatient?[ab]	
1st born	Yes/No/Don't know
2nd born	Yes/No/Don't know
Has either of the twins had to go back to hospital to stay overnight? (inpatient treatment)[ab]	
1st born	Yes/No/Don't know
2nd born	Yes/No/Don't know
Have either of your twins	
any medical problems in the first 2 weeks after birth?[ab]	
1st born	Yes/No/Don't know
2nd born	Yes/No/Don't know
physical problems (e.g., cleft lip, hole in the heart)?[ab]	
1st born	Yes/No/Don't know
2nd born	Yes/No/Don't know

Note. "Don't know's" were recoded as missing for all variables.
[a] Item included in Prenatal and Neonatal Problem scale.
[b] Item included in Neonatal Problems scale.
[c] This item was not included in the analyses due to very low frequency of response.

Child Development, January/February 2010, Volume 81, Number 1, Pages 183–199

Frontal Electroencephalogram Asymmetry, Salivary Cortisol, and Internalizing Behavior Problems in Young Adults Who Were Born at Extremely Low Birth Weight

Louis A. Schmidt, Vladimir Miskovic, Michael Boyle, and Saroj Saigal
McMaster University

The authors examined internalizing behavior problems at middle childhood, adolescence, and young adulthood and brain-based measures of stress vulnerability in 154 right-handed, nonimpaired young adults (*M* age = 23 years): 71 (30 males, 41 females) born at extremely low birth weight (ELBW; < 1,000 g) and 83 (35 males, 48 females) controls born at normal birth weight (NBW). Internalizing behavior problems increased from adolescence to young adulthood among ELBW individuals. ELBW adults exhibited greater relative right frontal electroencephalogram activity at rest and more concurrent internalizing behavior problems than NBW controls. Being born at ELBW may have subtle influences on brain–behavior relations even in survivors without major impairments and evidence of these influences may not emerge until young adulthood.

Extremely low birth weight (ELBW) is a phenomenon that provides developmental researchers with a model for understanding how adverse events early in life may impact brain development and developmental outcomes across a variety of domains. Infants born at ELBW (< 1,000 g) are the smallest and most at-risk infants. Because the survival rate of infants born at ELBW has markedly improved over the last few decades due to advances in neonatal care (see Lorenz, 2001, for a review), there is an increased need to track development in ELBW survivors.

Previous research from our group and others has established that children born at ELBW are at risk for a variety of neurodevelopmental and psychiatric disorders and stress-related problems. Saigal, Rosenbaum, Szatmari, and Campbell (1991) reported that the prevalence rate for a number of neurological and neurosensory impairments (e.g., cerebral palsy, hydrocephaly, microcephaly, blindness, and deafness) is significantly higher among children born at ELBW than in normal birth weight (NBW) matched controls. ELBW survivors are also at higher risk for a range of academic, behavioral, social, and psychiatric problems during preschool (Szatmari, Saigal, Rosenbaum, Campbell, & King, 1990; Taylor, Hack, & Klein, 1998), middle childhood (Miller, Bowen, Gibson, Hand, & Ungerer, 2001; Szatmari, Saigal, Rosenbaum, & Campbell, 1993; Taylor, Klein, & Hack, 2000), adolescence (Grunau, Whitfield, & Fay, 2004; Saigal, Hoult, Streiner, Stoskopf, & Rosenbaum, 2000; Saigal, Lambert, Russ, & Hoult, 2002; Saigal, Pinelli, Hoult, Kim, & Boyle, 2003), and young adulthood (Schmidt, Miskovic, Boyle, & Saigal, 2008) compared with their NBW counterparts.

A similar pattern of stress-related difficulties such as psychiatric, behavioral, and socioemotional problems noted in children, adolescents, and young adults born at ELBW has been documented in children, adolescents, and young adults who were born at very low birth weight (VLBW < 1,500 g; see Chapieski & Evankovich, 1997; Hack, Klein, & Taylor, 1995; Schothorst & van Engeland, 1996; Zwicker & Harris, 2008, for reviews). These

---

This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Social Science and Humanities Research Council of Canada (SSHRC) awarded to Louis Schmidt; a predoctoral fellowship from NSERC awarded to Vladimir Miskovic under the direction of Louis Schmidt; and grants from Canadian Institutes of Health and Research (CIHR; Grant MOP42536) and National Institute of Child Health and Human Development (NICHD; Grant RO1HD40219) awarded to Saroj Saigal and her colleagues. We wish to thank the participants for their continued cooperation; Lindsay Bennett, Sylvia Nowakowski, Caroline Parkin, and Diane Santesso for their help with data collection; Lorraine Hoult and Barbara Stoskopf for arranging the assessments; and Alison Niccols for helpful comments on earlier drafts.

Correspondence concerning this article should be addressed to Louis A. Schmidt, Department of Psychology, Neuroscience & Behaviour, McMaster University, Hamilton, ON L8S 4K1, Canada. Electronic mail may be sent to schmidtl@mcmaster.ca.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8101-0012

behaviors include: lower levels of social competence (Ross, Lipper, & Auld, 1990), more psychiatric and behavioral problems (Botting, Powls, Cooke, & Marlow, 1997; Cooke, 2004; Dahl et al., 2006; Hack, Cartar, Schluchter, Klein, & Forrest, 2007; Hack et al., 2002; Hack et al., 2004; Hack et al., 2005; Indredavik, Vik, Heyerdahl, Kulseng, & Brubakk, 2005; Levy-Schiff et al., 1994; Reijneveld et al., 2006; Richards, Kelly, Doyle, & Callanan, 2001; Sykes et al., 1997), and lower self-concept (Richards et al., 2001) in children, adolescents, and young adults born at VLBW. Hack et al. (2007) also found that, although VLBW participants reported similar health, well-being, and functioning compared with NBW controls, they exhibited greater risk avoidance. Hack et al. suggested that reduced resilience in VLBW individuals may prove detrimental to their future adult health. In other high-risk birth groups, Allin et al. (2006) recently reported significantly lower extraversion and higher neuroticism scores in adults born very preterm (< 33 weeks gestation age) compared with full-term control adults.

Although pre- and perinatal risk factors such as low birth weight and prematurity are presumed to shape the developing brain and predict the development of stress-related problems (Als, 1995), we know little about their impact on brain-based and behavioral measures of stress vulnerability in adult survivors. The reason for this lack of attention is largely that there have been relatively few opportunities to study ELBW survivors as young adults. Because medical advances have improved survival rates further for infants born at ELBW only in the last few decades (Lorenz, 2001), the long-term impact of pre- and perinatal risk factors on stress vulnerability can now be examined in adult survivors.

Frontal Electroencephalogram (EEG) Asymmetry and Stress Vulnerability

One approach to the study of stress vulnerability has been to use measures of EEG activity to examine patterns of regional brain asymmetry in infants, children, adults, and clinical populations (Davidson, 1993, 2000; Fox, 1991, 1994). Davidson (1993, 2000) and Fox (1991, 1994) have argued that emotions are organized around approach-withdrawal motivational tendencies. The experience of positive emotions (e.g., joy, happiness, interest) facilitates approach-related responses whereas the experience of negative emotions (e.g., fear, sadness, disgust) facilitates withdrawal-related responses. Moreover, the anterior cerebral hemispheres are differentially lateralized for approach–withdrawal responses and the regulation of emotion. The left frontal region subserves approach-related behaviors and emotions, whereas the right frontal region subserves withdrawal-related emotions.

Over the past three decades, researchers have used the frontal EEG asymmetry-emotion model as a theoretical platform to derive predictions about emotional regulation and socioemotional development (see Coan & Allen, 2004, for a review). Overall, studies have shown that the pattern of resting frontal EEG alpha asymmetry (a) reflects a bias to experience either positive or negative emotion, (b) is related to individual differences in affective style, and (c) predicts psychopathology across development. Individuals who exhibit greater relative left frontal EEG activity at rest are social, extraverted, and outgoing, whereas those who exhibit greater relative right frontal EEG activity at rest are shy, anxious, and depressed.

For example, infants who are fearful and easily distressed (Buss et al., 2003; Fox, Bell, & Jones, 1992), toddlers who are behaviorally inhibited (Calkins, Fox, & Marshall, 1996; Fox, Bell, & Calkins, 1994), and children who are socially withdrawn (Fox, Henderson, Rubin, Calkins, & Schmidt, 2001; Fox et al., 1995), shy (Theall-Honey & Schmidt, 2006) and score high on internalizing behaviors (Fox, Schmidt, Calkins, Rubin, & Coplan, 1996) all exhibit greater relative right frontal EEG activity at rest. Adults who are shy (Schmidt, 1999) and socially anxious (Beaton et al., 2008) are known to display greater relative right frontal EEG activity at rest. In addition, clinically depressed adults are characterized by greater relative right frontal EEG activity at rest even when their symptoms of depression are in remission (Henriques & Davidson, 1990). Individuals who exhibit stable resting right, versus left, frontal EEG activity across time are also known to score high on measures of internalizing behaviors such as behavioral inhibition and withdrawal (Sutton & Davidson, 1997).

Studies have noted short- and long-term stability over time and context in resting frontal EEG alpha asymmetry and power measures in healthy individuals as well as some clinical populations. Schmidt and his colleagues reported stability in frontal EEG asymmetry measured second-by-second in healthy infants (Schmidt, 2008) and across different stages of sleep in adults (Schmidt, Cote, Santesso, & Milner, 2003). Tomarken, Davidson, Wheeler, and Kinney (1992) found stability in frontal EEG asymmetry and power across 3 weeks in healthy adults. Still other studies have noted stability across a longer period of time (i.e., 8- and 16-week intervals) in frontal alpha asymmetry in nonmedicated clinically depressed individuals, where changes in asymmetry were not related to changes in clinical state (Allen, Urry, Hitt, & Coan, 2004). More recent studies by Fox's group have reported long-term stability in the resting frontal asymmetry metric over a 6-month to 3-year interval in preschool- (3–5 years) and school-aged (6–9 years) children (Vuga, Fox, Cohn, Kovacs, & George, 2008), and over a 1- to 3-year interval in healthy adults and in adults with a history of major depression (Vuga et al., 2006). More recently, Schmidt and colleagues have noted long-term stability in the pattern of frontal EEG alpha asymmetry over a 3-year period in a sample of stable community outpatients with schizophrenia before and after controlling for positive and negative symptoms of the disorder (Jetha, Schmidt, & Goldberg, 2009) and over a 6-month period in adolescent females exposed to child maltreatment (Milkovic, Schmidt, Georgiades, Boyle, & MacMillen, 2009). Because the pattern of frontal EEG asymmetry at rest is stable across time (Tomarken et al., 1992) and context (Schmidt et al., 2003), and its appearance early in life is predictive of socioemotional and stress-related problems (see Coan & Allen, 2004; Fox et al., 1995; Fox et al., 1996; Fox et al., 2001), some have argued that this metric indexes stress vulnerability (Coan & Allen, 2004; Davidson, 1993, 2000; Fox, 1991, 1994).

There have been, however, no studies that have examined the impact of pre- and perinatal risk factors on resting frontal EEG asymmetry measures in adult survivors. The EEG studies that do exist are concerned primarily with premature infants at birth in neonatal intensive care units (Victor, Appleton, Beirne, Marson, & Weindling, 2005) or those early in the postnatal period (Beckwith & Parmelee, 1986; Duffy, Als, & McNulty, 1990; Hayakawa et al., 2001; Vecchierini, Andre, & d'Allest, 2007) in relation to behavioral and cognitive outcomes. Some researchers also have examined the EEGs of premature infants in relation to later manifestations of congenital brain disorders (Ellingson, Dutch, & McIntire, 1974). A primary goal of the present study was to examine the patterns of resting frontal EEG alpha asymmetry in young adults born at ELBW.

Salivary Cortisol and Stress Vulnerability

A second approach to the study of stress vulnerability has been to use neuroendocrine measures such as salivary cortisol (see Gunnar, 2000; Gunnar & Tagle, 2008, for reviews). The endocrine system has long been implicated in the regulation of emotion (see Gunnar, 1994). Cortisol is one of the critical molecules of the endocrine system. It is produced in response to stress and is easily measured in saliva. A number of studies have shown that cortisol increases in response to stress and the experience of fear in human infants, children, and adults, and baseline salivary cortisol predicts individual differences in stress vulnerability (see Gunnar & Tagle, 2008, for a review).

Human infants with high basal cortisol levels exhibit more freezing behavior in response to novel stimuli and greater relative right frontal EEG activity (Buss et al., 2003). Infants who remained temperamentally distressed from 9 to 13 months of age exhibited increases in adrenocortical activity during the presentation of laboratory vignettes (Gunnar, Mangelsdorf, Larson, & Hertsgaard, 1989). Kagan, Reznick, and Snidman (1987, 1988) reported that temperamentally shy and behaviorally inhibited children exhibited higher morning basal cortisol levels compared with uninhibited children. Schmidt and his colleagues replicated the morning basal cortisol finding in temperamentally shy preschoolers (Schmidt et al., 1997). In addition, adults with agitated depression (i.e., depression with comorbid anxiety) are also known to be characterized by high basal cortisol levels (Gold, Drevets, & Charney, 2002; Gold, Goodwin, & Chrousos, 1988a, 1988b).

Gunnar and her colleagues found that stressful events early in life shaped neuroendocrine responses. Gunnar's group noted high cortisol levels among 6- to 12-year-old children who spent more than 8 months in Romanian orphanages in their first years of postnatal life compared with early adopted children and nonadopted North American children (Gunnar, Morison, Chisholm, & Schuder, 2001). Gunnar and her colleagues also reported that behaviorally inhibited children who were in insecure attachment relationships displayed greater increases in salivary cortisol during fear-eliciting conditions compared with uninhibited and securely attached children (Nachmias, Gunnar, Mangelsdorf, Parritz, & Buss, 1996).

There have been relatively few studies that have examined the impact of pre- and perinatal risk factors on adult salivary cortisol levels (although see Kajantie & Raikkonen, in press, for a recent review). One exception is the work of Wust, Entringer, Federenko, Schlotz, and Hellhammer (2005) who recently reported that birth weight was associated with salivary cortisol responses to psychosocial stress in adult life. Lower birth weight due to twin births predicted higher cortisol reactivity in young adulthood. A second goal of the present study was to examine patterns of resting baseline salivary cortisol levels in young adults born at ELBW.

The Present Study

We examined whether being born at ELBW would have any long-term negative impact on brain–behavior relations and stress vulnerability. More specifically, we were interested in testing whether ELBW survivors without major impairments would exhibit more internalizing behavior problems in middle childhood, adolescence, and young adulthood and more stress vulnerability as indexed by two brain-based measures than their NBW peers. To our knowledge, there appear to be no studies examining multiple brain-based and behavioral measures of stress vulnerability in young adult survivors who were born at ELBW. It is possible that there may be subtle influences on brain development due to pre- and perinatal risk factors that may only manifest later in life (see, e.g., Miskovich, Schmidt, Boyle, & Seigel, 2008). For example, a recent study reported reduced fertility in adult survivors of extreme prematurity (Swamy, Østbye, & Skjaerven, 2008). Although causality could not be inferred from this study, the results do raise the possibility that pre- and perinatal risk factors may negatively impact development across different domains in subtle ways that do not manifest (or are not observed) until young adulthood.

By design, we selected relatively homogenous groups from the ELBW and NBW cohorts composed of participants who were free from neurosensory or psychiatric impairment, right-handed, and tested at approximately the same time of day, as these variables are known to influence the study measures. The two groups did not differ significantly on key maternal and familial risk factors such as parental social class, education, family functioning, maternal psychopathology, and social support measures collected in middle childhood, adolescence, and young adulthood. The two groups also did not differ on any demographic, individual, or familial risk factors other than birth weight from infancy.

We collected regional resting brain EEG activity, baseline salivary cortisol, and measures of internalizing behavior problems in ELBW young adults and NBW controls. These measures were examined in relation to measures of child, adolescent, and young adult internalizing behavior problems. The ELBW cohort has been followed longitudinally from birth and assessed at key developmental milestones (see Saigal, Szatmari, Rosenbaum, Campbell, & King, 1990; Saigal et al., 1991, 1996, 2003; Saigal, Stoskopf, Streiner, Boyle, et al., 2006 Szatmari et al., 1990; Szatmari et al., 1993).

Hypotheses

We made the following predictions: (a) young adults born at ELBW would exhibit significantly more internalizing behavior problems, greater relative right frontal EEG activity at rest, and higher baseline salivary cortisol levels compared with their NBW controls; (b) internalizing behavior problems would increase from adolescence to young adulthood in ELBW adults; and (c) greater relative right frontal EEG activity would be related to concurrent internalizing behavior problems and higher salivary cortisol levels.

Method

Participants and Study Overview

Extremely low birth weight. The ELBW cohort comprised 397 predominantly Caucasian infants (birth weight = 501–1,000 g) born between 1977 and 1982 to residents of a geographically defined region in central-west Ontario, Canada, and recruited at birth. Of these 397 infants, 179 (45%) survived to hospital discharge from the neonatal intensive care unit. There were 13 late deaths; therefore, 166 survived to young adulthood. Adults who were born at ELBW were followed longitudinally from birth, and neurosensory impairments (e.g., cerebral palsy, blindness, deafness, mental retardation, microcephaly) were diagnosed by a neonatologist and a developmental pediatrician. By design, we chose to constitute a fairly homogeneous group of participants and therefore excluded individuals with neurosensory impairments ($n = 46$), psychiatric problems ($n = 2$), and those who were non-right-handed (cf. Oldfield, 1971; $n = 25$). We selected only right-handed individuals because left-handers are known to differ in the lateralization of emotion (Heller & Levy, 1981). Of the 93 ELBW adult survivors eligible to participate, 7 were lost to follow-up, 4 refused the entire young adulthood study, and 11 either refused or were unable to be recalled to the laboratory. A total of 71 (76%) eligible ELBW adult survivors participated in this study.

Normal birth weight. The NBW control group comprised 145 individuals who were selected from a random sample of children born at term and obtained from class lists of 8-year-old children from several local public school boards, and were of similar race, sex, and socioeconomic status distribution as the ELBW cohort (Szatmari et al., 1993). The controls were not followed prospectively from birth. Of the 145 NBW controls originally enrolled, we excluded those with neurosensory impairments (*n* = 3), non-right-handed individuals (*n* = 27), and those who had epilepsy and multiple sclerosis at young adulthood (*n* = 2). Of the 113 eligible NBW control participants, 4 were lost to follow-up, 7 refused the entire young adulthood study, and 19 either refused or were unable to be recalled to the laboratory. A total of 83 (74%) NBW adults participated in the present study (see Table 1 for an overview).

The 154 (ELBW, *n* = 71; NBW, *n* = 83) participants in this study were part of a larger study examining transition to young adulthood (Saigal, Stoskopf, Pinelli, et al., 2006; Saigal, Stoskopf, Streiner, Boyle, et al., 2006), health status (Saigal et al., 2007), and quality of life (Saigal, Stoskopf, Streiner, Paneth, et al., 2006) of individuals born at ELBW. The analyses next focus on the measures of internalizing behavior problems, resting regional EEG activity and salivary cortisol collected during their young adulthood visit to our laboratory as well as measures of internalizing behavior problems collected during their middle childhood and adolescent follow-up visits.

Table 1
Extremely Low Birth Weight (ELBW) and Normal Birth Weight (NBW) Sample Overview

ELBW	NBW
179 survivors of neonatal intensive care unit	145 enrolled at age 8 (no deaths)
⇓	
13 late deaths	
⇓	
166 survivors to young adulthood	
Exclusions	**Exclusions**
46 neurosensory impaired	3 neurosensory impaired
25 nonright handed	27 nonright handed
2 mental illness at young adulthood	2 epilepsy and multiple sclerosis at young adulthood
Eligible to participate	**Eligible to participate**
⇓	⇓
93	**113**
7 lost to follow-up	4 lost to follow-up
4 refused entire young adult study	7 refused entire young adult study
11 refused/unable to recall to lab	19 refused/unable to recall to lab
⇓	⇓
71 Participated (76%)	**83 Participated (74%)**

Instruments and Measures

Internalizing behavior problems measured at middle childhood. Both cohorts of participants and their parents (ELBW: *n* = 64; NBW: *n* = 80) were asked to complete the Ontario Child Health Study (OCHS; see Boyle et al., 1987) at the middle childhood visit (see Szatmari et al., 1993). The OCHS scales are composed of a psychiatric checklist that measures conduct disorder, oppositional defiant disorder (ODD), attention deficit hyperactivity disorder (ADHD), overanxious disorder, separation anxiety, and depression based on criteria of the *Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Revised* (*DSM–III–R*; American Psychiatric Association, 1987). The externalizing problems scale is formed by summing the first three scales and the internalizing problems by summing the last three scales. The Child Behavior Checklist (CBCL; Achenbach & Edelbrock, 1983) and the original OCHS scale (Boyle et al., 1987; Offord et al., 1987) provided the basic pools of items and additional items were added by consensus when necessary, to describe a particular disorder. Each item has three response categories *never or not true, somewhat or sometimes true*, or *very or often true*, with scores of 0, 1, or 2, respectively. Respondents were asked to base their answer on the preceding 6 months. Of particular interest to the present study was in the internalizing behavior problems scale (i.e., overanxious disorder, separation anxiety, and depression). All three subscales were interrelated within age (*ps* < .05). Accordingly, the three individual subscales were first *z*-scored and then summed to form a separate composite measure of *middle childhood internalizing behavior problems*. The composite measure of middle childhood internalizing behavior problems was then *z*-scored.

Internalizing behavior problems measured at adolescence. The participants in both cohorts (ELBW: *n* = 65; NBW: *n* = 74) were asked to complete the Ontario Child Health Study–Revised (OCHS–R; see Boyle et al., 1993) scales, which were self-administered at the adolescent visit (see Saigal et al., 2003). Of particular interest from the OCHS–R was the internalizing behavior problems scale (i.e., overanxious disorder + separation anxiety + depression). All three subscales were interrelated within age (*ps* < .05). Accordingly, the three individual subscales were first *z*-scored and then summed to form a separate composite measure of *adolescent internalizing behavior problems*. The composite measure of middle childhood internalizing behavior problems was then *z*-scored.

Internalizing behavior problems measured at young adulthood. At the young adult visit, the participants in both cohorts (ELBW: *n* = 71; NBW: *n* = 82) completed (a) the CBCL Youth Self-Report (YSR; Achenbach, 1991) measure of particular interest from the YSR was the internalizing behavior problems scale (social withdrawal + somatic complaints + anxiety/depression scales); (b) the Spielberg Trait Anxiety Inventory (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983), which is a 20-item Likert-format measure that indexes trait anxiety (e.g., "I am a steady person," "I lack self-confidence"); and (c) items related to social phobia from the interviewer administered World Organization Composite International Diagnostic Interview Short-Form (CIDI–SF; Kessler, Andrews, Mroczek, Ustun, & Wittchen, 1998). Sample items include: Have you ever had such a strong, unreasonable fear of: (a) giving a speech or speaking in public, (b) eating or drinking where someone could watch you, (c) talking to people because you might have nothing to say or might sound foolish, (d) writing while someone watches, (e) taking part or speaking in a meeting or class, (f) going to a party or other social outing? Response format is *yes* or *no*. The CBCL YSR Internalizing Scale, Spielberger Trait Anxiety, and CIDI–SF Social Phobia scale were all highly interrelated (*ps* < .05). Accordingly, each was first *z*-scored and then added together to form a composite measure of *young adulthood internalizing behavior problems*. This composite measure was then *z*-scored.

Demographic risk. Information on demographic factors was obtained from the parents through parent report questionnaires at the middle childhood (see Szatmari et al., 1993), adolescent (see Saigal et al., 2003), and young adult visits. Parental socioeconomic status was measured using the Hollingshead index of social position (Hollingshead, 1969). Maternal education was assessed by responses to the OCHS questionnaire (Boyle et al., 1987).

Maternal and familial risk. Information on family factors was obtained from parent report questionnaires at the middle childhood (see Szatmari et al., 1993), adolescent (see Saigal et al., 2003), and young adulthood visits. Maternal mood was measured by a 10-item scale derived from the Bradburn Scale of Positive and Negative affect (Bradburn, 1969). This scale is composed of 5 items that measure positive affect and 5 items that measure negative affect. Maternal depression was indexed using the Center for Epidemiologic Studies–Depression Scale (CES–D; Carpenter et al., 1998; Radloff, 1977). Maternal anxiety was measured by the Speilberger State–Trait Anxiety Inventory (Spielberger et al., 1983). Maternal social support was assessed using Social Support Index (SSI; McCubbin, Patterson, & Glynn, 1987). The SSI measures degree of social networks and connectedness. Family functioning was measured by the 12-item general functioning scale derived from the McMaster Family Assessment Device (Byles, Byrne, Boyle, & Offord, 1988). This instrument assesses family functioning on six dimensions: problem solving, communication, roles, affective responsiveness, affective involvement, and behavior control.

Individual risk. Young adults from both cohorts completed the SSI (McCubbin et al., 1987), the CES–D (Carpenter et al., 1998; Radloff, 1977), and the Rosenberg Self-Esteem Short Form (Rosenberg, 1965). Body mass index was measured by dividing participant weight by height, squared. Level of respiratory support (0 = *none* to 4 = *high*) at birth also was examined within the ELBW group.

Procedures

Upon arrival to the laboratory at the young adult visit, we explained the procedures to the participants and obtained written informed consent. After allowing the participants 30 min to adjust to the laboratory, three saliva samples were collected approximately 15 min apart. Then the participant was led to a room used for EEG testing. Regional EEG was recorded during a 2-min resting baseline (1 min eyes open [EO], 1 min eyes closed [EC]) condition. Studies have shown that EEG power estimates as short as 2 min have proven reliable (Allen et al., 2004; Schmidt, 2008).

After the EEG recording, the EEG cap was removed, and the participant was debriefed and provided with remuneration for participating. All procedures were conducted under the supervision of trained research staff and approved by the McMaster University Health Sciences Research Ethics Board. The study was conducted from 2001 to 2004.

EEG Procedures and Measures

EEG recording. Regional EEG was collected using a Lycra stretch cap. The cap electrodes are positioned according to the 10/20 system of the International Federation (Jasper, 1958). The experimenter used the blunt end of a Q-tip in combination with an abrasive gel (Omni-prep, D.O. Weaver & Co., Aurora, CO) and gently abraded each electrode surface. Each electrode site was then filled with a small amount of electrolyte gel that served as a conductor. Electrode impedances below 10 K ohms per site and within 500 ohms between homologous sites were considered acceptable.

EEG was recorded from the left and right anterior and posterior regions of the scalp (i.e., midfrontal: F3, F4; central: C3, C4). EEG activity was collected in posterior regions to examine if asymmetry differences were specific to the frontal region. All active EEG sites were referenced to Cz during acquisition.

A calibration signal (10 Hz/0.47 V rms sine wave) was input through each amplifier prior to each data collection. The output of this signal was 50 μV, with a gain of 10,000. The five channels were amplified by individual SA Instrumentation Bioamplifiers, with filter settings for all channels set at 1 Hz (high pass) and 100 Hz (low pass). The data from all five channels were digitized online at a sampling rate of 512 Hz.

EEG data reduction and quantification. The EEG data were visually scanned for artifact due to movement (e.g., eye blinks, body movements) and edited out using software developed by James Long Company (EEG Analysis Program, Caroga Lake, NY). If artifact was present in one channel, then data in all channels were excluded. All artifact-free EEG data were analyzed using a discrete Fourier transform, with a Hanning window of one sec width and 50% overlap. Regional EEG power (in μV^2) was derived in the alpha (8 to 13 Hz) frequency band separately for the EO and EC conditions. It is well documented that this EEG frequency band is linked to individual differences in stress vulnerability (see Coan & Allen, 2004; Davidson, 2000). Because EEG power in the EO and EC conditions was highly related for each of the sites ($rs > .80$, $ps < .01$), a composite measure of resting EEG alpha power was computed separately for each EEG site by averaging power in the EO and EC conditions. This aggregate measure is known to produce a more reliable estimate of EEG power and asymmetry than separate EO and EC conditions (see Tomarken et al., 1992). A separate EEG asymmetry measure was then computed for each region (e.g., F4 alpha power *minus* F3 alpha power). Because EEG power is inversely related to activity, negative values on the frontal asymmetry metric reflect greater relative right EEG activity (Davidson & Tomarken, 1989). The two study groups did not differ in the amount of EEG artifact data removed or in the amount of artifact-free EEG data used in the analyses.

Salivary Cortisol Procedures and Measures

Data collection. Each participant was asked to refrain from eating or drinking 1 hr prior to coming to the laboratory and to report any atypical recent life events, as these factors are known to influence cortisol levels. Each participant was asked to donate three individual saliva samples during the laboratory visit. Passive saliva was collected by having each participant expectorate into a 50-μl cryogenic tube. Approximately 50 μl of saliva were collected at each sample. The average time of day for saliva collection was approximately 1,300 hr. The time of day that the three saliva samples were collected did not vary between groups or systematically with any of the other study measures.

Salivary cortisol assay determinations. All saliva samples were transported on ice and stored at −20°C prior to assays. Saliva was centrifuged at 3,000 × *g* for 15 min, and only the supernatant was assayed in the Neuroendocriniology Laboratory at Brock University (St. Catharines, ON, Canada). All enzyme immunoassays were carried out on NUNC Maxisorb plates. Cortisol antibodies (R4866) and corresponding horseradish peroxidase conjugate were obtained from C. Munro of the Clinical Endocrinology Laboratory, University of California, Davis. Steroid standards were obtained from Steraloids, Inc. (Newport, RI). Each sample was assayed in duplicate and averages were used. Interplate variation (CV) was 6.45%, whereas intraplate variation was 6.51%. Because all three salivary cortisol samples were correlated ($ps < .05$), they were aggregated and an average measure of baseline salivary cortisol (expressed in ng/ml) was computed. The natural log (ln) of this composite measure was then computed to reduce skewness.

Missing Data

Of the 154 participants, EEG data were missing from 21 participants (6 ELBW, 15 NBW), owing to several reasons: 6 (3 ELBW, 3 NBW) refused the EEG procedure, 10 (2 ELBW, 8 NBW) were excluded either due to equipment failure ($n = 4$) or excessive movement and motor artifacts rendered their EEG unusable ($n = 6$), and 5 (2 ELBW, 3 NBW) were excluded because they had extreme EEG power values +3 *SD* above the mean. The

ELBW and NBW groups did not differ, nor did males and females disproportionately differ in the number of participants excluded or without usable data.

Results

Preliminary Analyses

Sociodemographic risk factors. Sociodemographic data for the ELBW and NBW young adult groups are provided in Table 2. Separate chi-square analyses revealed that there were no significant differences in group membership for gender, parental social class, and educational status (*ps* $> .05$). A between-groups t test also revealed that the two groups did not differ on age at assessment at the young adult visit ($p > .05$). In addition, there were no significant between-group differences on any of these sociodemographic factors at their middle childhood (see Szatmari et al., 1993) and adolescent (see Saigal et al., 2003) visits.

Table 2

Sociodemographic Variables for the Extremely Low Birth Weight (ELBW) and Normal Birth Weight (NBW) Groups

	Group		
Measure	ELBW ($n = 71$)	NBW ($n = 83$)	
Birth weight (in g)			
M	874	3,395	$t = -44.04$; $p < .0001$
SD	11	470	
Gestational age (in weeks)			
Maximum	34	Term	
Minimum	23		
M	27.59		
SD	2.18		
Sex (F/M)	41/30	48/35	$X^2 = 0.001$; $p > .05$
Parental social class[a]			$X^2 = 2.39$; $p > .05$
I–II	28	48	
III	23	15	
IV–V	17	18	
Current highest education (%)[b]			$X^2 = 2.97$; $p > .05$
< High school	7	5	
Completed high school	36	39	
Partial postsecondary	23	22	
Completed university/college	5	16	
Age at adult assessment (in years)			
M	23.3	23.6	$t = -1.44$; $p > .05$
SD	1.4	1.1	

[a] Based on Hollingshead (1969) classification. Data were missing for 3 ELBW participants and 2 NBW participants. [b] Data were missing for 1 NBW participant.

Individual, maternal, and familial risk factors. A separate analysis of variance (ANOVA) with group (ELBW, NBW) as the between-groups factor was performed on the separate individual, maternal, and familial risk factors. These analyses revealed that the two groups did not significantly differ on individual social support, depression, self-esteem, and body mass index, or on maternal social support, mood, depression, anxiety, social support, and family functioning measured at the young adult visit (*ps* $> .05$). In addition, the two groups did not differ on these individual, maternal, and familial risk factors at their middle childhood (see Szatmari et al., 1993) or adolescent (see Saigal et al., 2003) visits.

Between-Group Analyses

Internalizing behavior problems at middle childhood, adolescence, and young adulthood. We predicted that individuals born at ELBW would exhibit significantly more internalizing behavior problems than those born at NBW. To test this prediction, we performed a repeated measures analysis of variance (ANOVA) with group (ELBW, NBW) and sex (male, female) as the between-groups factors, and age (middle childhood, adolescence, young adulthood) as a within-groups factor. The dependent measure was the composite z-score measure of internalizing behavior problems.

The analyses revealed a significant main effect for sex, $F(1, 129) = 4.81$, $p < .05$. Overall, females ($M = 0.114$, $SE = 0.08$) exhibited significantly more internalizing behavior problems than males ($M = -0.153$, $SE = 0.09$). The analyses also revealed a significant Group × Age interaction, $F(2, 128) = 4.74$, $p < .01$ (see Figure 1).

To decompose this interaction, we performed a separate ANOVA at each age, with group (ELBW, NBW) as the between-groups factor. The dependent measure was the composite z-score measure of internalizing behavior problems at each age. As predicted, the analyses revealed a significant main effect for group on internalizing behavior problems only at young adulthood, $F(1, 131) = 4.97$, $p < .05$, Cohen's $d = 0.35$ (see Figure 1). ELBW adults ($M = 0.21$, $SE = 0.14$) exhibited significantly more internalizing behavior problems as young adults than their NWB peers ($M = -0.19$, $SE = 0.11$). Pairwise t tests also revealed that, as predicted, ELBW

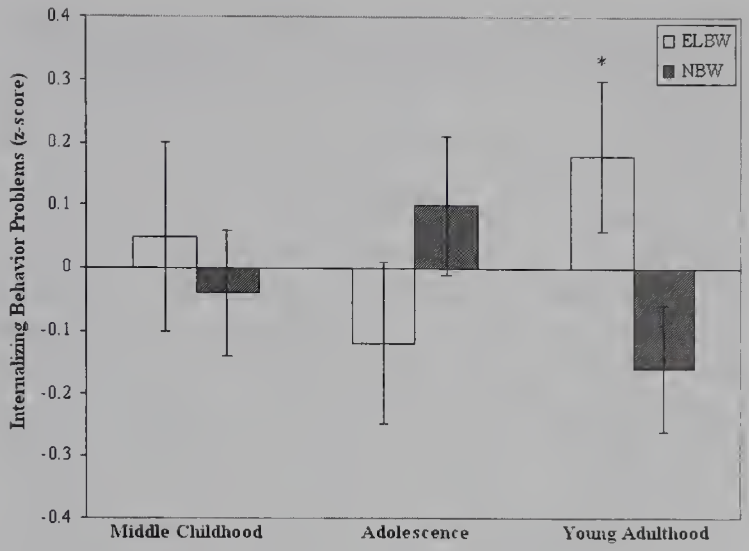


Figure 1. Standardized means (standard error bars) of internalizing problems in individuals born at extremely low birth weight (ELBW) and at normal birth weight (NBW) as children (ELBW: $n = 64$; NBW: $n = 80$) adolescents (ELBW: $n = 65$; NBW: $n = 74$), and young adults (ELBW: $n = 71$; NBW: $n = 82$; *between group difference, $p < .05$).

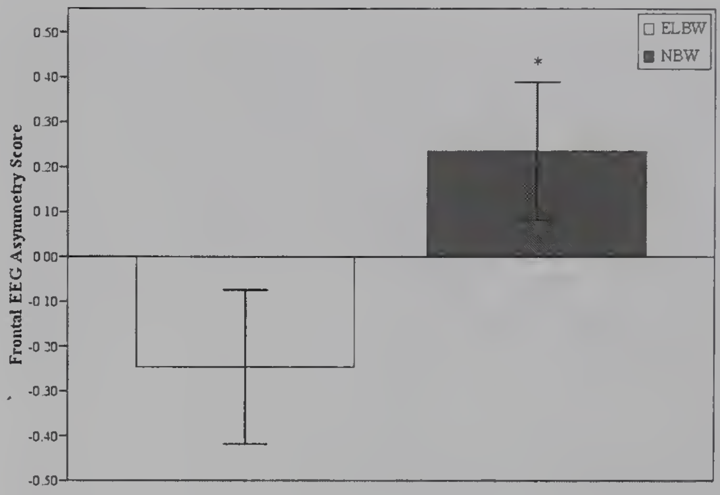


Figure 2. Mean difference (standard error bars) between individuals born at extremely low birth weight (ELBW; $n = 65$) and at normal birth weight (NBW; $n = 68$) on the pattern of resting frontal EEG alpha (8 to 13 Hz) asymmetry score (i.e., right frontal power *minus* left frontal power).

Note. EEG power is inversely related to activity so negative scores reflect greater relative right frontal EEG activity (*between group difference, $p < .05$).

adults showed a significant increase in internalizing behavior problems from adolescence ($M = -0.11$, $SE = 0.13$) to young adulthood ($M = 0.21$, $SE = 0.14$), $t(60) = 2.19$, $p < .05$, Cohen's $d = 0.28$, whereas NBW adults showed a significant decrease in internalizing behavior problems from adolescence ($M = 0.11$, $SE = 0.11$) to young adulthood ($M = -0.19$, $SE = 0.11$), $t(71) = 2.25$, $p < .05$, Cohen's $d = 0.29$. There were no significant between-group differences on internalizing behavior problems at middle childhood ($p > .05$) or adolescence ($p > .05$).

Resting frontal brain activity (EEG) at young adulthood. We predicted that adults born at ELBW would exhibit significantly greater relative right frontal EEG activity at rest compared with adults born at NBW. To test this prediction, we computed a frontal EEG asymmetry (right frontal power – left frontal power) metric. An analysis of covariance (ANCOVA) with group (ELBW, NBW) as the between-groups factor and concurrent depression ratings (CES–D measure) as a covariate was then performed on the frontal EEG asymmetry metric. Concurrent depression was covaried because concurrent mood is known to influence frontal asymmetry measures (see Beaton et al., 2008; Henriques & Davidson, 1990). This analysis revealed a significant main effect for group on resting frontal EEG asymmetry measure before, $F(1, 131) = 4.39$, $p < .05$, Cohen's $d = 0.37$, and after $F(1, 129) = 3.95$, $p = .05$, Cohen's $d = 0.36$, controlling for concurrent depression ratings (see Figure 2). As predicted, adults born at ELBW ($M = -0.25$, $SE = 0.17$) exhibited significantly greater relative right frontal EEG activity at rest compared with adults born at NBW ($M = 0.24$, $SE = 0.17$).

To understand the source of the asymmetry difference, a repeated measures ANOVA was computed with group (ELBW, NBW) and sex (male, female) as the between-groups factors, and region (frontal, central) and hemisphere (left, right) as the within-groups factors. The dependent measure was EEG alpha (8 to 13 Hz) power (μV^2). There were no significant main or interaction effects for sex, so this variable was collapsed in all subsequent analyses. The analyses revealed a significant Group × Region × Hemisphere interaction, $F(1, 131) = 4.59$, $p < .05$.

To decompose this interaction, we performed a separate ANOVA for each region with group (ELBW, NBW) as the between-subjects factor and hemisphere (left, right) as the within-subjects factor. The dependent measure was EEG alpha (8 to 13 Hz) power (μV^2). The analyses revealed a significant Group × Hemisphere interaction on alpha power only in the frontal region, $F(1, 131) = 4.39$, $p < .05$, confirming the frontal asymmetry difference above. The ANOVA with group and hemisphere as between-subjects factors on EEG power for the central region was not significant, ($p > .05$). Follow-up between-group and pairwise t tests computed on left and right frontal and central EEG power failed to yield any significant differences ($ps > .05$).

Baseline salivary cortisol at young adulthood. Table 3 presents the descriptive statistics on the salivary

Table 3

Descriptive Statistics for Extremely Low Birth Weight (ELBW) and Normal Birth Weight (NBW) Young Adults on the Salivary Cortisol Measures

	Group		
Measure	ELBW	NBW	
ln baseline salivary cortisol			
M (in ng/ml)	0.962	1.047	$t = -0.37$; $p > .05$
SE	0.150	0.171	
Average time of day of saliva collection			
M (in hr)	13:25	13:42	$t = -0.40$; $p > .05$
SE (in min)	0:29	0:17	

cortisol measures collected for the two groups. We predicted that ELBW adults would exhibit significantly higher salivary cortisol at baseline compared with adults born at NBW.

An ANCOVA with group (ELBW, NBW) and sex (male, female) as the between-groups factors was computed on the average baseline salivary cortisol measure. Time of day was included as a covariate in the salivary cortisol analysis, given that it is known to be related to cortisol levels (Daly & Evans, 1974). The analysis failed to reveal any significant main or interaction effects on the average natural log baseline salivary cortisol measure ($p > .05$).

Within-Group Analyses

Full sample analyses. Table 4 presents the Pearson correlations among the study measures for the full sample. As can be seen in Table 4, internalizing behavior problems at middle childhood and young adulthood ($r = .21$, $p < .01$) and at adolescence and young adulthood ($r = .33$, $p < .001$) were significantly correlated. As expected, children and adolescents who exhibited internalizing behavior problems were likely to display internalizing behavior problems as young adults.

Frontal EEG asymmetry at rest was related to concurrent internalizing behavior problems ($r = -.14$, $p < .06$) and baseline salivary cortisol ($r = -.20$, $p < .05$) at the young adult visit, although the former relation only approached statistical significance. As predicted, adults who exhibited greater relative right frontal EEG activity also tended to exhibit more internalizing behavior problems and higher baseline salivary cortisol levels contemporaneously.

Table 4

Pearson Correlations Among Study Measures Across Participants

Measure	1	2	3	4	5
1. Child internalizing problems	—				
2. Adolescent internalizing problems	.06	—			
3. Young adult internalizing problems	.21**	.33**	—		
4. Resting frontal EEG alpha asymmetry	.11	−.11	$-.14^{\dagger}$	—	
5. ln baseline salivary cortisol	−.01	.02	.04	−.20*	—

Note. Time of day was partialled out of all correlations involving the ln baseline salivary cortisol measure; EEG power is inversely related to activity so negative scores reflect greater relative right frontal EEG activity.

$^{\dagger}p < .06$. $^{*}p < .05$. $^{**}p < .01$.

Within-ELBW group analyses. We computed a series of within-group Pearson correlations within the ELBW group to examine if the two measures of stress vulnerability (i.e., resting frontal EEG asymmetry and baseline salivary cortisol) covaried in any systematic way with adverse early life events (e.g., birth weight, gestational age and respiratory support). Table 5 presents the within-group Pearson correlations among the study measures for the ELBW group.

As can be seen in Table 5, resting frontal EEG asymmetry was related to baseline salivary cortisol ($r = -.23$, $p < .05$) and birth weight ($r = .22$, $p < .05$) but not gestation age ($r = .03$, $p > .05$) or level of respiratory support at birth ($r = -.08$, $p > .05$). As predicted, ELBW adults who exhibited greater relative right frontal EEG activity at rest were likely to

Table 5

Pearson Correlations Among Study Measures Within Participants in the Extremely Low Birth Weight Group

Measure	1	2	3	4	5
1. Birth weight	—				
2. Gestation age	.34*	—			
3. Respiratory support level at birth	−.39*	−.23	—		
4. Resting frontal EEG alpha asymmetry	.22*	.03	−.08	—	
5. ln baseline salivary cortisol	.01	.01	−.08	−.23*	—

Note. Time of day was partialled out of all correlations involving the ln baseline salivary cortisol measure; EEG power is inversely related to activity so negative scores reflect greater relative right frontal EEG activity.

$^{*}p < .05$.

exhibit higher salivary cortisol levels and lower birth weights.

Discussion

Pre- and perinatal risk factors such as being born at ELBW are presumed to influence brain–behavior relations. We examined internalizing behavior problems at middle childhood, adolescence, and young adulthood and patterns of resting frontal brain activity and baseline salivary cortisol in a sample of adult survivors of ELBW without major impairments. As predicted, we found that internalizing behavior problems significantly increased from adolescence to young adulthood among ELBW adults. Also as predicted, we found that ELBW young adults exhibited significantly greater relative right frontal EEG activity and more internalizing behavior problems than NBW controls. These findings were based, overall, on medium effect sizes and are consistent with a large corpus of literature suggesting a relation between right frontal brain asymmetry at rest and individual differences in stress vulnerability (see Coan & Allen, 2004; Davidson, 2000; Fox, 1991, for substantive reviews). Here, we extend these earlier findings on resting frontal EEG asymmetry as a stress vulnerability marker to a sample of adults born at ELBW without major impairments.

Why did differences in internalizing behaviors emerge in young adulthood and not earlier in development in ELBW adults? It is possible that there may be subtle influences on brain development due to adverse events early in life that may only manifest later in life. A recent study reported reduced fertility in adult survivors of extreme prematurity (Swamy et al., 2008). It is possible that adverse life events such as being born at ELBW may negatively impact brain–behavior relations among ELBW survivors only later in life at developmental transition points. One such developmental transition is young adulthood, which is characterized by psychosocial stressors such as the development of personal autonomy, striving for financial independence, career development, romantic relationships, and starting a family. This speculation needs to be tempered, however, in that a recent study noted pervasive behavior problems at 6 years of age in a sample of children born extremely premature (i.e., ≤ 25; Samara, Marlow, & Wolke, 2008), so behavioral problems may exist in extremely premature populations before adulthood. However, it is also important to note that the study by Samara et al. (2008) included a heterogenous sample of premature children likely including children with impairments. Such impairments may have contributed to their behavioral problems at younger ages. In the present study, we used a nonimpaired sample of adult survivors by design to mitigate some potential confounds due to impairment.

We also found that right frontal EEG asymmetry predicted concurrent internalizing behavior problems and cortisol levels across the full sample, and was significantly related to higher cortisol levels and lower birth weights among ELBW adults. These results are consistent with findings from studies of stress responses in humans (Buss et al., 2003) and nonhuman primates (Kalin, Larson, Shelton, & Davidson, 1998). Buss et al. (2003) found that human infants with high basal cortisol levels exhibited more freezing behavior in response to novel stimuli and greater relative right frontal EEG activity at age 6 months than infants with relatively lower levels of basal cortisol. Kalin et al. (1998) reported that nonhuman primate Rhesus monkeys who exhibited right frontal EEG activity also displayed high basal cortisol levels and fearful behavior. Taken together, these two studies suggest that (a) the pattern of resting right frontal EEG asymmetry is predictive of high baseline cortisol and stress responses, (b) this pattern emerges early in life, and (c) it plays a role in stress dysregulation that is conserved across mammals. The results of this study extend these findings to adult survivors of ELBW.

What does resting right frontal brain asymmetry and its relation to high cortisol reflect in ELBW? Over the last three decades, there have been more than 100 studies published that have examined patterns of frontal EEG activity during the processing of emotion and its role in explaining individual differences in affective style (see Coan & Allen, 2004, for a review). Overall, these studies suggest that individuals exhibit greater relative left frontal EEG during the processing of positive emotion (e.g., happy, joy), and greater relative right frontal EEG activity during the processing of negative emotion (e.g., fear, sadness). These patterns are evidenced across a variety of sensory modalities used to induce emotion, including visual, auditory, imagery, and olfactory senses. The findings from these studies also suggest that the pattern of resting frontal EEG activity emerges in the first months of postnatal life, remains moderately stable during development from childhood through adolescence and young adulthood, and predicts individual differences in affective style and risk for psychopathology. Individuals who present with tonic left frontal EEG activity tend to be social, extraverted, and outgoing, whereas individuals who exhibit tonic right frontal EEG activity are shy, anxious, and depressed.

It is possible that the pattern of greater relative right frontal EEG activity may constitute a mechanism predisposing individuals to difficulties in regulating stress, such as physical and mental problems. The pattern of right frontal EEG asymmetry observed in adult survivors of ELBW without major impairments might be due to adverse events early in life that shaped brain development in a relatively subtle way. Adverse early experiences are known to alter the development of the prefrontal brain regions, resulting in differences in right, but not left, adult prefrontal volumes in nonhuman primates (Lyons, Afarian, Schatzberg, Sawyer-Glover, & Moseley, 2002). The observed relation between right frontal EEG asymmetry and high cortisol in young adults born at ELBW may reflect a less obvious impairment due to this adverse event early in life, placing some young adults born at ELBW at risk for stress-related problems.

What may account for the lack of significant group differences on the salivary cortisol measure used in this study? Because the hypothalamic–pituitary–adrenal (HPA) axis is a dynamic system, it is possible that individuals, despite having similar basal salivary cortisol, may have different adrenocorticotrophic hormone (ACTH) levels. That is, although ELBW and NBW adults have similar cortisol levels (possibly for down-regulation of adrenal ACTH receptors), it is possible that ELBW participants have greater ACTH levels than controls. This hypothesis is in line with what others have reported in nondepressed adult survivors of childhood trauma (see Gunnar & Quevedo, 2007; Heim et al., 2000).

It also may be that adverse events early in life such as low birth weight only perturb the neuroendocrine system early in postnatal life on reactivity measures (see Wust et al., 2005) and not on basal measures. Or, by young adulthood, development and maturational control of this system may have been possibly repaired. The lack of group differences also might be related to issues of differential risk and resiliency and specificity of different stress vulnerability systems. The impact of ELBW may affect some, but not all, brain-based aspects of stress vulnerability in the same way. Some brain systems may be more vulnerable to the impact of adverse early life events than others.

Theoretical Implications

The findings from this study may be interpreted by, and have theoretical implications for, at least two models of development. One model of development is a brain "insult" hypothesis. It is possible that the events producing reduced birth weight also have perturbed select brain systems involved in the regulation of stress. Right frontal brain asymmetry and internalizing behavior problems may develop as a result of early brain insult in survivors without major impairments. We found that lower birth weight predicted right frontal brain asymmetry at young adulthood in ELBW survivors without impairments.

It is important to note, however, that there is an emerging literature suggesting another variation of a brain insult model. That is, the brain develops quite differently at extreme prematurity if the infant is raised extrauterine (e.g., cortical folding), resulting in the building of a different brain (e.g., Kapellou et al., 2006). This idea would predict that the impact is not specific to any one function, but influences general domains of functioning (e.g., language difficulties and school achievement) as has been shown for both very and extremely premature children (e.g., Wolke, Samara, Bracewell, & Marlow, 2008).

A second model that the present results may inform is a fetal programming hypothesis (Barker, 1995). Although controversial, Barker posited that stressful events during fetal development may "program" the developing brain for how it handles subsequent stress. A proposed mechanism is perturbations in the HPA axis due to early life events, which contribute to chronically elevated cortisol levels (Fall et al., 2002) and autonomic responses (Phillips & Jones, 2006). This early programming places individuals born extremely premature at risk for disease (Reynolds et al., 2001) and may influence individual differences in temperamental style (Pesonen et al., 2006). However, other studies have not found support for the fetal programming hypothesis. For example, Hanna et al. (1993) failed to find chronically elevated cortisol levels in infant survivors of ELBW. Hanna et al. argued that the relatively low levels of cortisol noted in infants born at ELBW may be due to the extremely premature brain's inability to recognize the stress of the illness or because of inadequate hypothalamic secretion of corticotropin-releasing hormone. If the fetal programming hypothesis were correct, we would have expected to find between group differences on salivary cortisol levels. As noted earlier, we did not find between-group differences on the cortisol measure. It might be that although there were no differences between the groups on cortisol, the differences might exist at the level of ACTH. However, we did not index ACTH, but future studies should consider doing so to test this hypothesis. The results of this study also suggest that stress early in life may alter fetal programming of specific stress vulnerability measures (e.g., resting frontal brain asymmetry), but not others (e.g., baseline salivary cortisol).

Limitations

There are at least four limitations in this study. First, we recorded resting frontal EEG asymmetry and baseline salivary cortisol levels on one occasion only. Accordingly, we do not have information on the stability of these measures or information on how frontal EEG asymmetry may or may not covary with internalizing behavior problems at earlier ages. Although empirical evidence suggests that the pattern of resting frontal brain activity is stable across time (Tomarken et al., 1992; Vuga et al., 2006; Vuga et al., 2008), it would be advisable to have repeated measures over time and earlier in postnatal development. Second, EEG is measured at the scalp and thought to reflect cortical electrical activity. However, we do not know the exact sources of the EEG signal so caution needs to be exercised in terms of interpretation of EEG findings as being specific to the cerebral cortex and a particular brain region. Future studies should consider the use of more dense arrays of EEG electrodes than used in the present study so that EEG source analyses can be performed. As well, measures of functional magnetic imaging might prove useful in combination with electrocortical measures to provide accurate spatial and temporal resolution of brain-based measures of stress vulnerability. Third, issues of causation cannot be ascertained. Although we presume that the adverse early life event of being born ELBW preceded the development of right frontal EEG asymmetry, we need to caution against making any kind of causal interpretation. It might be that right frontal EEG asymmetry is a consequence of ELBW, or a coexisting factor. Tracking brain development at pre- and early postnatal stages may hold some of the answers to questions of causality. Fourth, although we studied adult survivors of ELBW, there are other populations characterized by early adversity that are of interest. For example, survivors of VLBW (Hack et al., 2002), babies born small for their gestational age (Strauss, 2000), and babies born prematurely (Schothorst, Swaab-Barneveld, & van Engeland, 2007) are also known to be at risk for stress-related problems. In future studies examining multiple brain-based measures of stress vulnerability, these populations would further inform our understanding of the relation between early adversity and stress vulnerability.

Conclusions

Pre- and perinatal risk factors such as being born at ELBW are known to have a profound impact on a range of developmental processes. We found that internalizing behavior problems significantly increased from adolescence to young adulthood among ELBW individuals without major impairments. We also found that ELBW adults exhibited significantly more internalizing behavior problems and greater relative right frontal EEG activity at rest than NBW controls. Greater relative right frontal EEG activity was related to higher cortisol levels and lower birth weights among ELBW adults. Findings suggest that ELBW may negatively influence brain–behavior relations in subtle ways even among survivors without major impairments and that evidence of these influences may not emerge until young adulthood. ELBW may influence the development of right frontal brain asymmetry, a putative mechanism for difficulties regulating stress responses, which can manifest as internalizing behavior problems in young adulthood.

References

Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist*. Burlington: University of Vermont, Department of Psychiatry.

Achenbach, T. M., & Edelbrock, C. (1983). *Manual for the Child Behavior Checklist and Revised Behavioral Profile*. Burlington: University of Vermont, Department of Psychiatry.

Allen, J. J., Urry, H. L., Hitt, S. K., & Coan, J. A. (2004). The stability of resting frontal electroencephalographic asymmetry in depression. *Psychophysiology, 41*, 269–280.

Allin, M., Rooney, M., Cuddy, M., Wyatt, J., Walshe, M., Rifkin, L., et al. (2006). Personality in young adults who are born preterm. *Pediatrics, 117*, 309–316.

Als, H. (1995). The preterm infant: A model for the study of fetal brain expectation. In J. P. Lecanuet, W. P. Fifer, N. A. Krasnegor, & W. P. Smotherman (Eds.), *Fetal development: A psychobiological perspective* (pp. 439–472). Hillsdale, NJ: Erlbaum.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.

Barker, D. J. P. (1995). Fetal origins of coronary heart disease. *British Medical Journal, 311*, 171–174.

Beaton, E. A., Schmidt, L. A., Ashbaugh, A. R., Santesso, D. L., Antony, M. M., & McCabe, R. E. (2008). Resting and reactive frontal brain electrical activity (EEG) among a non-clinical sample of socially anxious adults: Does concurrent depressive mood matter? *Neuropsychiatric Disease and Treatment, 4*, 187–192.

Beckwith, L., & Parmelee, A. H., Jr. (1986). EEG patterns of preterm infants, home environment, and later IQ. *Child Development, 57*, 777–789.

Botting, N., Powls, A., Cooke, R. W. I., & Marlow, N. (1997). Attention deficit hyperactivity disorders and other psychiatric outcomes in very low birth weight children at 12 years. *Journal of Child Psychology and Psychiatry, 38*, 931–941.

Boyle, M. H., Offord, D. R., Hofmann, H. G., Catlin, G. P., Byles, J. A., Cadman, D. T., et al. (1987). Ontario Child Health Study: Methodology. *Archives of General Psychiatry, 44*, 826–831.

Boyle, M. H., Offord, D. R., Racine, Y., Fleming, J. E., Szatmari, P., & Sanford, M. (1993). Evaluation of the revised Ontario Child Health Study Scales. *Journal of Child Psychology and Psychiatry, 34*, 189–213.

Bradburn, N. M. (1969). *The structure of psychological wellbeing*. Chicago: Aldine.

Buss, K. A., Schumacher, J. R., Dolski, I., Kalin, N. H., Goldsmith, H. H., & Davidson, R. J. (2003). Right frontal brain activity, cortisol, and withdrawal behavior in 6-month-old infants. *Behavioral Neuroscience, 117*, 11–20.

Byles, J., Byrne, C., Boyle, M. H., & Offord, D. R. (1988). Ontario Child Health Study: Reliability and validity of the general functioning subscale of the McMaster Family Assessment Device. *Family Process, 27*, 97–104.

Calkins, S. D., Fox, N. A., & Marshall, T. R. (1996). Behavioral and physiological antecedents of inhibited and uninhibited behavior. *Child Development, 67*, 523–540.

Carpenter, J. S., Andykowski, M. A., Wilson, J., Hall, L. A., Rayens, M. K., Sachs, B., et al. (1998). Psychometrics for two short forms of the Center for Epidemiologic Studies–Depression Scale. *Issues Mental Health Nursing, 19*, 481–494.

Chapieski, M. L., & Evankovich, K. D. (1997). Behavioral effects of prematurity. *Seminar in Perinatology, 21*, 221–239.

Coan, J. A., & Allen, J. J. B. (2004). Frontal EEG asymmetry as a moderator and mediator of emotion. *Biological Psychology, 67*, 7–49.

Cooke, R. W. (2004). Health, lifestyle, and quality of life in young adults born very preterm. *Archives of Disease in Childhood, 89*, 201–206.

Dahl, L. B., Kaaresen, P. I., Tunby, J., Handegard, B. H., Kvernmo, S., & Ronning, J. A. (2006). Emotional, behavioral, social, and academic outcomes in adolescents born with very low birth weight. *Pediatrics, 118*, e449–e459.

Daly, J. R., & Evans, J. I. (1974). Daily rhythms of steroid and associated pituitary hormones in man and their relationship to sleep. In M. H. Biggs & G. A. Christie (Eds.), *Advances in steroid biochemistry and pharmacology* (pp. 61–109). New York: Academic Press.

Davidson, R. J. (1993). The neuropsychology of emotion and affective style. In M. Lewis & J. M. Haviland (Eds.), *Handbook of emotion* (pp. 143–154). New York: Guilford.

Davidson, R. J. (2000). Affective style, psychopathology, and resilience: Brain mechanisms and plasticity. *American Psychologist, 55*, 1196–1214.

Davidson, R. J., & Tomarken, A. J. (1989). Laterality and emotion: An electrophysiological approach. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (pp. 419–441). New York: Elsevier.

Duffy, F. H., Als, H., & McNulty, G. B. (1990). Behavioral and electrophysiological evidence for gestation age effects in healthy preterm and fullterm infants studied two weeks after expected due date. *Child Development, 61*, 1271–1286.

Ellingson, R. J., Dutch, S. J., & McIntire, M. S. (1974). EEG's of prematures: 3–8 year follow-up study. *Developmental Psychobiology, 7*, 529–538.

Fall, C. H. D., Dennison, E., Cooper, C., Pringle, J., Kellinggray, S. D., & Hindmarsh, P. (2002). Does birth weight predict adult serum cortisol concentrations? Twenty-four hour profiles in the United Kingdom 1920–1930 Hertfordshire Birth Cohort. *Journal of Clinical Endocrinology & Metabolism, 87*, 2001–2007.

Fox, N. A. (1991). If it's not left, it's right: Electroencephalograph asymmetry and the development of emotion. *American Psychologist, 46*, 863–872.

Fox, N. A. (1994). Dynamic cerebral processes underlying emotion regulation. In N. A. Fox (Ed.), The development of emotion regulation: Behavioral and biological considerations. *Monographs of the Society for Research in Child Development, 59*(2–3, Serial No. 240).

Fox, N. A., Bell, M. A., & Calkins, S. D. (1994). Neural plasticity and development in the first two years of life: Evidence from cognitive and socioemotional domains of research. *Development and Psychopathology, 6*, 677–696.

Fox, N. A., Bell, M. A., & Jones, N. A. (1992). Individual differences in response to stress and cerebral asymmetry. *Developmental Neuropsychology, 8*, 161–184.

Fox, N. A., Henderson, H. A., Rubin, K. H., Calkins, S. D., & Schmidt, L. A. (2001). Continuity and discontinuity of behavioral inhibition and exuberance: Psychophysiological and behavioral influences across the first four years of life. *Child Development, 72*, 1–21.

Fox, N. A., Rubin, K. H., Calkins, S. D., Marshall, T. R., Coplan, R. J., Porges, S. W., et al. (1995). Frontal activation asymmetry and social competence at four years of age. *Child Development, 66*, 1770–1784.

Fox, N. A., Schmidt, L. A., Calkins, S. D., Rubin, K. H., & Coplan, R. J. (1996). The role of frontal activation in the regulation and dysregulation of social behavior during the preschool years. *Development and Psychopathology, 8*, 89–102.

Gold, P. W., Drevets, W. C., & Charney, D. S. (2002). New insights into the role of cortisol and the glucocorticoid receptor in severe depression. *Biological Psychiatry, 52*, 381–385.

Gold, P. W., Goodwin, F. K., & Chrousos, G. P. (1988a). Clinical and biochemical manifestations of depression (Part I). *New England Journal of Medicine, 319*, 348–353.

Gold, P. W., Goodwin, F. K., & Chrousos, G. P. (1988b). Clinical and biochemical manifestations of depression (Part II). *New England Journal of Medicine, 319*, 413–420.

Grunau, R. E., Whitfield, M. F., & Fay, T. B. (2004). Psychosocial and academic characteristics of extremely low birth weight (≤800 g) adolescents who are free of major impairment compared with term-born control subjects. *Pediatrics, 114*, e725–e732.

Gunnar, M. R. (1994). Psychoendocrine studies of temperament and stress in early childhood: Expanding current models. In J. E. Bates & T. D. Wachs (Eds.), *Temperament: Individual differences at the interface of biology and behavior* (pp. 175–198). Washington, DC: American Psychological Association.

Gunnar, M. R. (2000). Early adversity and the development of stress reactivity and regulation. In C. A. Nelson (Ed.), *The effects of adversity on neurobehavioral development: Minnesota symposium on child psychology* (Vol. 31, pp. 163–200). Mahwah, NJ: Erlbaum.

Gunnar, M. R., Mangelsdorf, S., Larson, M., & Hertsgaard, L. (1989). Attachment, temperament, and adrenocortical activity in infancy: A study of psychoendocrine regulation. *Developmental Psychology, 25*, 355–363.

Gunnar, M. R., Morison, S. J., Chisholm, K., & Schuder, M. (2001). Salivary cortisol levels in children adopted from Romanian orphanages. *Development and Psychopathology, 13*, 611–628.

Gunnar, M., & Quevedo, K. (2007). The neurobiology of stress and development. *Annual Review of Psychology, 58*, 145–173.

Gunnar, M. R., & Tagle, N. M. (2008). Neuroendocrine measures in developmental research. In L. A. Schmidt & S. J. Segalowitz (Eds.), *Developmental psychophysiology: Theory, systems, and methods* (pp. 343–364). New York: Cambridge University Press.

Hack, M., Cartar, L., Schluchter, M., Klein, N., & Forrest, C. B. (2007). Self-perceived health, functioning and well-being of very low birth weight infants at age 20 years. *Journal of Pediatrics, 151*, 635–641.

Hack, M., Flannery, D., Schluchter, M., Cartar, L., Borawski, E., & Klein, N. (2002). Outcomes in young adulthood for very-low-birth-weight infants. *New England Journal of Medicine, 346*, 149–157.

Hack, M., Klein, N. K., & Taylor, H. G. (1995). Long-term developmental outcomes of low birth weight infants. *The Future of Children, 5*, 176–196.

Hack, M., Youngstrom, E. A., Cartar, L., Schluchter, M., Taylor, H. G., Flannery, D., et al. (2004). Behavioral outcomes and evidence of psychopathology among very low birth weight infants at age 20 years. *Pediatrics, 114*, 932–940.

Hack, M., Youngstrom, E. A., Cartar, L., Schluchter, M., Taylor, G. H., Flannery, D. J., et al. (2005). Predictors of internalizing symptoms among very low birth weight young women. *Journal of Developmental & Behavioral Pediatrics, 26*, 93–104.

Hanna, C. E., Keith, L. D., Colasurdo, M. A., Buffkin, D. C., Laird, M. R., Mandel, S. H., et al. (1993). Hypothalamic pituitary adrenal function in the extreme low birth weight infant. *Journal of Clinical Endocrinology and Metabolism, 76*, 384–387.

Hayakawa, M., Okumura, A., Hayakawa, F., Watanabe, K., Ohshiro, M., Kato, Y., et al. (2001). Background electroencephalographic (EEG) activities of very preterm infants born at less than 27 weeks gestation: A study on the degree of continuity. *Archives of Disease in Childhood, Fetal Neonatology Edition, 84*, F163–F167.

Heim, C., Newport, J., Heit, S., Graham, Y. P., Wilcox, M., Bonsall, R., et al. (2000). Pituitary-adrenal and autonomic responses to stress in women after sexual and physical abuse in childhood. *Journal of the American Medical Association, 284*, 592–597.

Heller, W., & Levy, J. (1981). Perception and expression of emotion in right-handers and left-handers. *Neuropsychologia, 19*, 263–272.

Henriques, J. B., & Davidson, R. J. (1990). Regional brain electrical asymmetries discriminate between previously depressed and healthy controls. *Journal of Abnormal Psychology, 99*, 22–31.

Hollingshead, A. B. (1969). *Two-factor index of social position (mimeograph).* New Haven, CT: Yale University Press.

Indredavik, M. S., Vik, T., Heyerdahl, S., Kulseng, S., & Brubakk, A. M. (2005). Psychiatric symptoms in low birth weight adolescents, assessed by screening questionnaires. *European Child & Adolescent Psychiatry, 14*, 226–236.

Jasper, H. H. (1958). The ten-twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology, 10*, 371–375.

Jetha, M. K., Schmidt, L. A., & Goldberg, J. O. (2009). Long-term stability of resting frontal EEG alpha asymmetry and power in a sample of stable community outpatients with schizophrenia. *International Journal of Psychophysiology, 72*, 228–233.

Kagan, J., Reznick, J. S., & Snidman, N. (1987). The physiology and psychology of behavioral inhibition in children. *Child Development, 58*, 1459–1473.

Kagan, J., Reznick, J. S., & Snidman, N. (1988). Biological basis of childhood shyness. *Science, 240*, 167–171.

Kajantie, E., & Raikkonen, K. (in press). Early life predictors of the physiological stress response later in life. *Neuroscience and Biobehavioral Reviews.*

Kalin, N. H., Larson, C., Shelton, S. E., & Davidson, R. J. (1998). Asymmetric frontal brain activity, cortisol, and behavior associated with fearful temperament in Rhesus monkeys. *Behavioral Neuroscience, 112*, 286–292.

Kapellou, O., Counsell, S. J., Kennea, N., Dyet, L., Saeed, N., Stark, J., et al. (2006). Abnormal cortical development after premature birth shown by altered allometric scaling of brain growth. *PLoS Medicine, 3*, e265.

Kessler, R. C., Andrews, G., Mroczek, D., Ustun, B., & Wittchen, H. U. (1998). The World Health Organization Composite International Diagnostic Interview Short-Form (CIDI–SF). *International Journal of Methods Psychiatric Research, 7*, 171–185.

Levy-Schiff, R., Einat, G., Har-Even, D., Mogilner, M., Mogilner, S., Lerman, M., et al. (1994). Emotional and behavioral adjustment in children born prematurely. *Journal of Clinical Child Psychology, 23*, 323–333.

Lorenz, J. M. (2001). The outcome of extreme prematurity. *Seminars in Perinatology, 25*, 348–359.

Lyons, D. M., Afarian, H., Schatzberg, A. F., Sawyer-Glover, A., & Moseley, M. E. (2002). Experience-dependent asymmetric variation in primate prefrontal morphology. *Behavioural Brain Research, 136*, 51–59.

McCubbin, H., Patterson, J., & Glynn, T. (1987). Social Support Index (SSI). In H. McCubbin & A. Thompson (Eds.), *Family assessment inventories for research and practice* (pp. 283–302). Madison: University of Wisconsin–Madison.

Miller, M., Bowen, J. R., Gibson, F. L., Hand, P. J., & Ungerer, J. A. (2001). Behaviour problems in extremely low birthweight children at 5 and 8 years of age. *Child: Care, Health and Development, 27*, 569–581.

Miskovic, V., Schmidt, L. A., Boyle, M., & Saigal, S. (2009). Regional electroencephalogram (EEG) spectral power and hemispheric coherence in young adults born at extremely low birth weight. *Clinical Neurophysiology, 120*, 231–238.

Miskovic, V., Schmidt, L. A., Georgiades, K., Boyle, M., & MacMillan, H. L. (2009). Stability of resting frontal electroencephalogram (EEG) asymmetry and cardiac vagal tone in adolescent females exposed to child maltreatment. *Developmental Psychobiology, 51*, 474–487.

Nachmias, M., Gunnar, M., Mangelsdorf, S., Parritz, R., & Buss, K. A. (1996). Behavioral inhibition and stress reactivity: The moderating role of attachment security. *Child Development, 67*, 508–522.

Offord, D. R., Boyle, M. H., Szatmari, P., Rae-Grant, N. I., Links, P. S., Cadman, D. T., et al. (1987). Ontario Child Health Study II: Six-month prevalence of disorder and rates of service utilization. *Archives of General Psychiatry, 44*, 832–836.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia, 9*, 97–113.

Pesonen, A. K., Raikkonen, K., Kajantie, E., Heinonen, K., Strandberg, T. E., & Jarvenpaa, A. L. (2006). Fetal programming of temperamental negative affectivity among children born healthy at term. *Developmental Psychobiology, 48*, 633–643.

Phillips, D. I., & Jones, A. (2006). Fetal programming of autonomic and HPA function: Do people who were small babies have enhanced stress responses? *Journal of Physiology, 572*, 45–50.

Radloff, L. (1977). The CES–D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 3*, 385–401.

Reijneveld, S. A., de Kleine, M. J. K., van Baar, A. L., Kollee, L. A. A., Verhaak, C. M., Verhulst, F. C., et al. (2006). Behavioural and emotional problems in very preterm and very low birthweight infants at age 5 years. *Archives of Disease in Childhood, Fetal and Neonatal Edition, 91*, F423–F428.

Reynolds, R. M., Walker, B. R., Syddall, H. E., Andrew, R., Wood, P. J., Whorwood, C. B., et al. (2001). Altered control of cortisol secretion in adult men with low birth weight and cardiovascular risk factors. *Journal of Clinical Endocrinology & Metabolism, 86*, 245–250.

Richards, A. L., Kelly, E. A., Doyle, L. W., & Callanan, C. (2001). Cognition, academic progress, behavior and self-concept at 14 years of very low birth weight children. *Journal of Developmental and Behavioral Pediatrics*, 22, 11–18.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Ross, G., Lipper, E. G., & Auld, P. A. M. (1990). Social competence and behavior problems in premature children at school age. *Pediatrics, 86*, 391–397.

Saigal, S., Feeny, D., Rosenbaum, P., Furlong, W., Burrows, E., & Stoskopf, B. (1996). Self-perceived health status and health related quality of life in extremely low-birth-weight infants at adolescence. *Journal of the American Medical Association, 276*, 453–459.

Saigal, S., Hoult, L. A., Streiner, D. L., Stoskopf, B. L., & Rosenbaum, P. L. (2000). School difficulties at adolescence in a regional cohort of children who were extremely low birth weight. *Pediatrics, 105*, 325–331.

Saigal, S., Lambert, M., Russ, C., & Hoult, L. (2002). Self-esteem of adolescents who were born prematurely. *Pediatrics, 109*, 429–433.

Saigal, S., Pinelli, J., Hoult, L., Kim, M. M., & Boyle, M. (2003). Psychopathology and social competencies of adolescents who were extremely low birth weight. *Pediatrics, 111*, 969–975.

Saigal, S., Rosenbaum, P., Szatmari, P., & Campbell, D. (1991). Learning disabilities and school problems in a regional cohort of extremely low birthweight (<1000 g) children: A comparison with matched term controls. *Journal of Developmental and Behavioral Pediatrics*, 12, 294–300.

Saigal, S., Stoskopf, B., Boyle, M., Paneth, N., Pinelli, J., Streiner, D., et al. (2007). Comparison of current health, functional limitations and health care utilization of young adults born with extremely low birthweight and normal birthweight. *Pediatrics, 119*, e562–e573.

Saigal, S., Stoskopf, B., Pinelli, J., Streiner, D., Hoult, L., Paneth, N., et al. (2006). Self-perceived health-related quality of life of former extremely low birthweight infants at young adulthood. *Pediatrics, 118*, 1140–1148. (Erratum: Dr. Michael Boyle)

Saigal, S., Stoskopf, B., Streiner, D., Boyle, M., Pinelli, J., Paneth, N., et al. (2006). Transition of extremely low-birthweight infants from adolescence to young adulthood: Comparison with normal birth-weight controls. *Journal of the American Medical Association*, *295*, 667–675.

Saigal, S., Stoskopf, B., Streiner, D., Paneth, N., Pinelli, J., & Boyle, M. (2006). Growth trajectories of extremely low birthweight infants from birth to young adulthood: A longitudinal population-based study. *Pediatric Research*, *60*, 751–758.

Saigal, S., Szatmari, P., Rosenbaum, P., Campbell, D., & King, S. (1990). Intellectual and functional status at school entry of children who weighed 1000 grams or less at birth: A regional perspective of births in the 1980's. *Journal of Pediatrics*, *116*, 409–416.

Samara, M., Marlow, N., & Wolke, D. (2008). Pervasive behavior problems at 6 years of age in a total-population sample of children born at ≤25 weeks gestation. *Pediatrics*, *122*, 562–573.

Schmidt, L. A. (1999). Frontal brain electrical activity in shyness *and* sociability. *Psychological Science*, *10*, 316–320.

Schmidt, L. A. (2008). Patterns of second-by-second resting frontal brain (EEG) asymmetry and their relation to heart rate and temperament in 9-month-old human infants. *Personality and Individual Differences*, *44*, 216–225.

Schmidt, L. A., Cote, K. A., Santesso, D. L., & Milner, C. E. (2003). Frontal electroencephalogram alpha asymmetry during sleep: Stability and its relation to affective style. *Emotion*, *3*, 401–407.

Schmidt, L. A., Fox, N. A., Schulkin, J., & Gold, P. W. (1999). Behavioral and psychophysiological correlates of self-presentation in temperamentally shy children. *Developmental Psychobiology*, *35*, 119–135.

Schmidt, L. A., Miskovic, V., Boyle, M., & Saigal, S. (2008). Shyness and timidity in young adults who were born at extremely low birth weight. *Pediatrics*, *122*, e181–e187.

Schothorst, P. F., Swaab-Barneveld, H., & van Engeland, H. (2007). Psychiatric disorders and MND in non-handicapped preterm children: Prevalence and stability from school age into adolescence. *European Child & Adolescent Psychiatry*, *16*, 439–448.

Schothorst, P. F., & van Engeland, H. (1996). Long-term behavioral sequelae of prematurity. *Journal of the American Academy of Child and Adolescent Psychiatry*, *35*, 175–183.

Spielberger, C. D., Gorsuch, R. L., Lushene, P. R., Vagg, P. R., & Jacobs, A. G. (1983). *Manual for the State-Trait Anxiety Inventory (Form Y)*. Palo Alto, CA: Consulting Psychologists Press.

Strauss, R. S. (2000). Adult functional outcome of those born small for gestation age. *Journal of the American Medical Association*, *283*, 625–632.

Sutton, S. K., & Davidson, R. J. (1997). Prefrontal brain asymmetry: A biological substrate of the behavioral approach and inhibition systems. *Psychological Science*, *8*, 204–210.

Swamy, G. K., Østbye, T., & Skjaerven, R. (2008). Association of preterm birth with long-term survival, reproduction, and next-generation preterm birth. *Journal of the American Medical Association*, *299*, 1429–1436.

Sykes, D. H., Hoy, E. A., Bill, J. M., McClure, B. G., Halliday, H. L., & Reid, M. (1997). Behavioral adjustment in school of very low birthweight children. *Journal of Child Psychology and Psychiatry*, *38*, 315–325.

Szatmari, P., Saigal, S., Rosenbaum, P., & Campbell, D. (1993). Psychopathology and adaptive functioning among extremely low birthweight children at eight years of age. *Development and Psychopathology*, *5*, 345–357.

Szatmari, P., Saigal, S., Rosenbaum, P., Campbell, D., & King, S. (1990). Psychiatric disorders at age five years among children with birthweights < 1000 g: A regional perspective. *Developmental Medicine of Child Neurology*, *32*, 954–962.

Taylor, H. G., Hack, M., & Klein, N. (1998). Attention deficits in children with <750 gram birthweight. *Child Neuropsychology*, *4*, 21–34.

Taylor, H. G., Klein, N., & Hack, M. (2000). School-age consequences of birth weight less than 750 g: A review and update. *Developmental Neuropsychology*, *17*, 289–321.

Theall-Honey, L. A., & Schmidt, L. A. (2006). Do temperamentally shy children process emotion differently than non-shy children? Behavioral, psychophysiological, and gender differences in reticent preschoolers. *Developmental Psychobiology*, *48*, 187–196.

Tomarken, A. J., Davidson, R. J., Wheeler, R. E., & Kinney, L. (1992). Psychometric properties of resting anterior EEG asymmetry: Temporal stability and internal consistency. *Psychophysiology*, *29*, 576–592.

Vecchierini, M. F., Andre, M., & d'Allest, A. M. (2007). Normal EEG of premature infants born between 24 and 30 weeks gestation age: Terminology, definitions and maturation aspects. *Clinical Neurophysiology*, *37*, 311–323.

Victor, S., Appleton, R. E., Beirne, M., Marson, A. G., & Weindling, A. M. (2005). Spectral analysis of electroencephalography in premature newborn infants: Normal ranges. *Pediatric Research*, *57*, 336–341.

Vuga, M., Fox, N. A., Cohn, J. F., George, C. J., Levenstein, R. M., & Kovacs, M. (2006). Long-term stability of frontal electroencephalographic asymmetry in adults with a history of depression and controls. *International Journal of Psychophysiology*, *59*, 107–115.

Vuga, M., Fox, N. A., Cohn, J. F., Kovacs, M., & George, C. J. (2008). Long-term stability of electroencephaplographic asymmetry and power in 3 to 9 year old children. *International Journal of Psychophsyiology*, *67*, 70–77.

Wolke, D., Samara, M., Bracewell, M., & Marlow, N. (2008). Specific language difficulties and school achievement in children born at 25 weeks of gestation or less. *Journal of Pediatrics*, *152*, 256–262.

Wust, S., Entringer, S., Federenko, I. S., Schlotz, W., & Hellhammer, D. H. (2005). Birth weight is associated with salivary cortisol responses to psychosocial stress in adult life. *Psychoneuroendocrinology*, *30*, 591–598.

Zwicker, J. G., & Harris, S. R. (2008). Quality of life of formerly preterm and very low birth weight infants from preschool age to adulthood: A systematic review. *Pediatrics*, *121*, e366–e376.

Child Development, January/February 2010, Volume 81, Number 1, Pages 200–211

Late-Treated Phenylketonuria and Partial Reversibility of Intellectual Impairment

Scott D. Grosse
National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention

Individuals with late-treated phenylketonuria (PKU) not detected by newborn screening but who followed dietary treatment for at least 12 months before 7 years of age have intelligence quotient (IQ) scores that range from severe impairment to the low-normal range. Among adults with late-treated PKU in California, 85% of those who were born from 1961 to 1978 had IQ scores of 70 or above. Longitudinal studies with repeated cognitive assessments often show average changes in cognitive test scores as high as 20–45 points. Although the severe cognitive impairment associated with untreated PKU can in many cases be partially reversed with dietary treatment, prompt initiation of treatment following newborn metabolic screening is essential for optimal development and the prevention of disability.

Newborn screening for phenylketonuria (PKU), which began in the early 1960s (Levy & Albers, 2000), is a major public health success. PKU is an inborn error of metabolism caused by mutation in the phenylalanine (Phe) hydroxylase gene (*PAH*) gene resulting in elevated levels of Phe that can cause serious brain damage (Scriver, 2007).

The combination of early detection at birth and adherence to a low-Phe diet has been found to prevent all cases of intellectual disability among children born with PKU according to population-based studies that have linked newborn screening records with developmental disability surveillance systems (Bower, Leonard, & Petterson, 2000; Van Naarden Braun, Yeargin-Allsopp, Schendel, & Fernhoff, 2003).

It is commonly stated that almost all individuals with untreated classic PKU experience severe intellectual disability, typically with intelligence quotient (IQ) < 30 (Waisbren et al., 2007). However not all individuals with untreated PKU experience intellectual disability (Hanley, 2007), with certain individuals protected by the blood–brain barrier from high Phe levels passing from the bloodstream to the brain. Most, though, do experience some degree of impairment. One British study reported that 2 (15%) of 13 older siblings with untreated PKU among probands identified on the basis of developmental deficits had IQ scores > 70 and attended ordinary classes, but with markedly lower performance than their unaffected siblings (Smith & Wolff, 1974). A retrospective study from Australia of 56 individuals with classic PKU born prior to 1965 found that 75% of adults had IQ scores < 52 and 93% had IQ scores ≤ 85 (Ramus, Forrest, Pitt, & Cotton, 1999).

Treatment for PKU with a low-Phe diet that keeps blood Phe levels within the 2–6 mg/dl range prior to adolescence can eliminate all adverse effects on intellectual development (Stemerdink et al., 2000). However, it is very difficult to keep to a strict diet, and consequently, specific deficits, such as executive function, attention, and visual–spatial ability, remain problematic in many individuals with early and continuously treated PKU who have moderately elevated blood Phe levels (Brumm et al., 2004; Burgard, 2000; Diamond & Herzberg, 1996; Diamond, Prevor, Callender, & Druin, 1997).

The conventional wisdom holds that late-treated PKU results in "severe, irreversible mental

---

The findings and conclusions in this report are those of the author and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Helpful comments were received from Adele Diamond, Jaime Frias, Michael Glass, Muin Khoury, Loïc de Parscau, John Thompson, Edwin Trevathan, Susan Waisbren, Connie Whitehead, and two anonymous referees. Kumiko Imai and Lijing Ouyang provided translation assistance.

Correspondence concerning this article should be addressed to Scott D. Grosse, Centers for Disease Control and Prevention, 1600 Clifton Road, Mail Stop E-87, Atlanta, GA 30333. Electronic mail may be sent to sgrosse@cdc.gov.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8101-0013

retardation'' (Abadie et al., 2001; Feldman, 1994) and that dietary treatment ''does not reverse preexisting damage'' (Seymour et al., 1997). ''Late-treated'' PKU commonly refers to children who were not detected through screening and were put on a low-Phe diet after 2–3 months of age, However, older studies noted that initiation of dietary treatment was often associated with substantial improvements in test scores (Berman, Waisman, & Graham, 1966; Knox, 1972; Waisbren, Schnell, & Levy, 1980). More recently, research conducted in California conclusively established that long-term outcomes among adults with late-treated PKU were often not as dire as expected (Koch et al., 1999; Trefz, Cipcic-Schmidt, & Koch, 2000). In interpreting these findings, Levy (2000) acknowledged ''some degree of reversibility'' following the late initiation of treatment, which he stated should be ''encouraging to families who have to cope with this tragedy.'' This does not detract from the importance of screening and prompt treatment to avoid needless brain damage.

Research on the neuropathology of brain impairment caused by PKU has yielded valuable information as to the mechanisms and anatomical manifestations of brain damage resulting from PKU. It has long been known that high Phe levels in the bloodstream alone do not explain brain damage, because individuals with the same elevated blood Phe levels show marked variations in the extent of neurological damage (Gizewska et al., 2003; Möller, Weglage, Wiedermann, & Ullrich, 1998; Möller et al., 2003; Weglage et al., 2002). The important new evidence is that interindividual variations in the blood–brain barrier appear to determine brain Phe levels and hence neurological outcomes (Gizewska et al., 2003; Koch, Moats, Guttler, Guldberg, & Nelson, 2000).

Magnetic resonance imaging (MRI) of the brains of individuals with PKU has revealed the characteristic presence of white matter (WM) abnormalities or alterations associated with elevated Phe levels (Anderson et al., 2004; Pietz et al., 1996). It has been demonstrated that these WM changes are partially reversible with short-term dietary treatment (Cleary et al., 1995; Thompson et al., 1993; Ullrich et al., 1994). Variations in WM lesions have not been reported to be statistically significantly associated with variations in IQ (Leuzzi et al., 2007; Pérez-Dueñas et al., 2006; Ullrich et al., 1994), although both children and adults with moderate WM abnormalities have been reported to have IQ scores lower by 6–14 points on average than those with no detectable WM lesions (Anderson et al., 2007; Koch et al., 2002). Crucially, WM abnormalities have been linked to specific neurological deficits such as in executive functions (Anderson et al., 2004; Anderson et al., 2007; Brumm et al., 2004), which have also been demonstrated to result from reduced dopamine activity in the prefrontal cortex in treated PKU children with insufficiently strict diets (Diamond, 1996; Diamond, Ciaramitaro, Donner, Djali, & Robinson, 1994; Diamond et al., 1997). Recent research suggests that neurological damage in late-treated PKU appears to be highly correlated with differences in the volume and distribution of gray matter in the brain (Pérez-Dueñas et al., 2006).

This article addresses the extent to which reductions in cognitive ability associated with late-diagnosed PKU are partially reversible by dietary modification. First, we review cross-sectional data on cognitive test scores and the prevalence of intellectual disability of late-treated individuals with PKU in relation to the age at which a low-Phe diet was introduced. Second, data on longitudinal changes in cognitive test scores among late-treated individuals are reviewed. No attempt is made to review the voluminous literature documenting the cognitive effects of age of initiation and adherence to therapy among individuals with early-treated PKU (see Burgard, 2000; Waisbren et al., 2007).

Method

Terms and Measures

Phenylketonuria is a type of hyperphenylalanemia (HPA), which is characterized by persistently elevated concentrations of the amino acid Phe in the blood. Persons with HPA are classified as having either classic PKU (blood Phe > 1200 µmol/L or 20 mg/dl), mild PKU (blood Phe 600–1200 µmol/L or 10–20 mg/dl), or non-PKU HPA (blood Phe 120–599 µmol/L or 2–10 mg/dl; Scriver & Kaufman, 2001). Mutations on the *PAH* gene, which result in deficiency of activity of a liver enzyme, PAH, are responsible for most cases of each type of HPA (Kaye & the Committee on Genetics, American Academy of Pediatrics, 2006). In this article, PKU refers specifically to classic phenylketonuria.

A low-Phe diet is one that avoids most foods with natural protein and incorporates Phe-free formula and/or medical foods with low levels of Phe. The purpose of a low-Phe diet is to maintain a blood Phe concentration for children within a target range, set at 120–360 µmol/L (2–6 mg/dl) in the United States (Koch et al., 2000). Metabolic control should be achieved as early as possible within the 1st month of life.

The terminology relating to cognitive impairment and disability is in a state of flux. Traditionally, the terms *mental retardation* and *mental handicap* were used to characterize individuals with cognitive abilities below the normal range and with limitations in adaptive functioning that began during childhood. Alternative terms, *intellectual disability* in the United States and *learning disability* in the United Kingdom, are increasingly preferred. In this article, *intellectual disability* (ID) refers to individuals with IQ or developmental quotient (DQ) scores ≤ 70 who are presumed to experience disability, and it is used as a substitute for *mental retardation*. An IQ score of 70 (or 75) is commonly used as a cutoff point for ID; individuals with IQ scores of 50–70 are classified as having mild ID and those with lower scores are said to have moderate, severe, or profound ID (Roeleveld, Zielhuis, & Gabreels, 1997). IQ scores of 70–84 are often classified as borderline and scores of 85 or higher are considered normal.

The cognitive tests used to calculate IQ scores are typically administered to children only at 4 years of age or older, whereas tests administered to younger children yield DQ scores. DQ and IQ scores are both standardized on the same scale, but DQ scores at or before 12 months of age primarily measure performance on sensorimotor tasks (Slater, 1997). DQ scores assessed at 2 years of age or later include verbal assessments and are more predictive of later IQ scores, although differences in test types still complicate the interpretation of changes in scores (Hack et al., 2005; Slater, 1997). For example, among preterm infants who had DQ < 70 at 20 months of age, 20% of those without neurosensory impairments had IQ < 70 at 8 years of age, compared with 62% of those who had a specific impairment (Hack et al., 2005). The gains were not primarily an artifact of the use of different tests, because individual test scores at both points at time were predictive of academic skills at 8 years of age.

Comparisons of cognitive test scores over time, whether between cohorts or longitudinal assessments for individuals, can be affected by general increases over time in IQ scores (Flynn, 1999). IQ tests are periodically recalibrated or renormed every decade or so, but the tendency for mean IQ scores to drift up over time means that a 2 *SD* difference from the mean IQ score might be > 75. However, this effect would not account for differential increases in test scores between groups. Also, there is no evidence that the percentage of children classified as having intellectual disability based on low IQ scores has decreased over time (Bhasin, Brocksen, Avchen, & Van Naarden Braun, 2006).

Inclusion and Exclusion Criteria

A search was undertaken to identify studies that assessed cognitive outcomes or neurological manifestations among individuals with late-treated PKU. Only studies in which individuals began dietary treatment before 7 years of age were included. Also, children had to have followed a low-Phe diet for at least 12 months to be included. Multiple search strategies were used, including systematic searches in PubMed and searches of citation lists. No restriction was placed on the date of publication for articles in English, Spanish, French, or German, and articles published after 2000 were included regardless of language of publication. A PubMed search conducted on April 18, 2007, for "phenylketonuria" and "late treatment" identified 76 studies, most of which dealt with treatment during adulthood, maternal PKU, or discontinuation of a low-Phe diet during adolescence rather than late initiation of treatment during childhood; 24 articles from this search met the inclusion criteria. Another search for "phenylketonuria" and "IQ" resulted in 179 hits, which included three studies of late-treated PKU not identified in the earlier search. Several other relevant studies were identified by clicking on the "Related Articles" link in PubMed for studies meeting the inclusion criteria. Finally, other articles were identified through searches of citation lists in other publications, including review articles.

Additional exclusion criteria were imposed for cross-sectional studies, owing to the relatively large number of studies compared with only a handful of studies reporting longitudinal cognitive assessments for late-treated individuals. First, because DQ scores in early childhood are less reliable than IQ scores measured in school-age children (Waisbren et al., 2007), only cross-sectional studies reporting IQ scores in children over 4 years of age were retained in this section. Second, only studies from North America, Western Europe, or other high-income countries were retained in this section because of the often poor quality of dietary control in the past in low- and middle-income countries. The quality of dietary treatment has a mediating influence on neurological outcomes in PKU, and the question of interest for this review consists of outcomes for late-treated individuals who receive high-quality management once diagnosed. Similarly, because individuals who were put on low-Phe diets beginning in the late 1950s or early 1960s typically had poorer dietary control than those who were treated in later periods, it is important to compare outcomes for successive birth cohorts.

Cognitive Assessments in Late-Treated PKU

Cross-sectional assessments. Challenges in interpreting data on cognitive assessments in individuals with late-treated PKU include different types of cognitive tests, ascertainment bias, and the quality of treatment provided. Ascertainment bias can affect studies in which children were detected as having PKU on the basis of testing prompted by developmental deficits, thereby biasing cognitive scores downward relative to children detected through biochemical screening (Birch & Tizard, 1967). Ascertainment bias continues to characterize many of the studies of late-treated PKU discussed here. Other potential biases include differences among families that might be associated with children's developmental outcomes, the timing of diagnosis, and the level of adherence to dietary restriction. Few of the studies reviewed here attempted to control for such potential confounding.

Even the quality of dietary control, which is the most important influence on developmental outcomes, was only partially controlled for in most published tabulations. Few of the studies reviewed here reported the degree of dietary control in sufficient detail that would allow for comparison of findings across studies. The reader should keep in mind that reported cognitive scores in late-diagnosed children could be biased downward relative to early-treated children if late-diagnosed children on average have less tightly controlled diets.

Prior to the introduction of dietary treatment and newborn screening, most individuals with PKU had IQ scores < 50 (Ramus et al., 1999). After dietary treatment was introduced in the 1950s, test scores began to improve and mean scores were often reported to be between 50 and 60. For example, McBean and Stephenson (1968) reported that 61% of 23 late-treated children in Scotland had IQ < 60. Two studies from Canada reported mean IQ scores of approximately 55 for children treated beginning between 6 and 24 months of age. Scriver, Katz, and Clow (1968) reported that 43 such children in Quebec had a mean IQ of 56 when tested at a mean age of 5 years, and Hanley, Linsao, and Netley (1971) reported a mean IQ of 55 for 30 such children in Ontario. If one includes in the latter calculations 20 children diagnosed after 24 months of age, mean IQ was 49.4. A subsequent analysis reported that for 27 young adults, mean age 23 years, whose treatment had began at an average of 24 months of age and lasted for an average of 4.5 years, mean IQ was 57.6 (Netley, Hanley, & Rudner, 1984).

Similarly, a 1980 study from Hannover, Germany, reported information on 3 individuals up to 25 years of age who had been diagnosed and begun treatment for PKU at 3.5–8.5 years of age and 3 others who were never put on a restricted diet (Schweitzer, Wolff, & Brodehl, 1987). It appears that all had been born prior to newborn screening being introduced. All 6 had IQ scores < 70; all 3 untreated individuals required residential care for severe disability as did 1 of 3 late-treated individuals. The mean IQ for the latter 3 was 55. However, another German study, from Hamburg, which involved 8 late-treated children with PKU, reported graphically raw scores on the Columbia Mental Maturity Scale. The average raw score was about 35, which was below the normal range, but for the subset of late-treated children who had good dietary control the mean score was about 45, within the normal range and not significantly lower than for early-treated children with good dietary control (Steinhausen & Börner, 1977).

A study from Japan reported data on 5 adults 24–33 years of age with late-detected PKU who participated in a MRI research study. All had been diagnosed at 1.5–2.75 years of age and subsequently put on a low-Phe diet (Izumi, Yamazaki, Nakabayashi, & Owada, 2006). One subject was reported to have severe mental retardation and 3 others were said to have mild or borderline mental retardation, but no criteria were given for these categories or IQ test results reported.

Other studies from high-income countries reported average IQ scores of approximately 70 in late-treated children born prior to the introduction of newborn screening. Among 5 children in Minnesota, United States, who were put on a low-Phe diet at 6–27 months of age, mean IQ at 7–11 years of age was 77 (Siegel, Balow, Fisch, & Anderson, 1968). Two larger studies yielded more representative data. First, Dobson et al. (1968) reported that 64 U.S. and English children who were placed on a low-Phe diet at 4–36 months of age and assessed at age 7 years had mean IQ of 69, which did not vary by age of initiation of treatment. Second, Smith et al. (1978) reported that 43 late-treated children in England and Germany who had been placed on a strict diet for at least 3 years after 4 months of age had mean IQ of 71 when assessed at a mean of 12 years of age.

In the United States, Legido et al. (1993) reported IQ results for 12 individuals with late-treated PKU seen at a Philadelphia clinic, all except one of whom had been born prior to newborn screening. All had been started on a low-Phe diet after age 3 months and had gone off the diet at a mean age of 7 years. Although all had been identified as having PKU as part of a workup for developmental delay, mean IQ in adolescence was 72.8, and 4 (33%) had final IQ < 70. Age of diagnosis was reported to be a significant influence on final IQ, which was 80.8 for 5 who had been put on a low-Phe diet before 10 months of age and 67.1 for 7 whose treatment started later. The authors did not comment on the contrast between their finding of an effect of age of treatment and the null finding of Dobson et al. (1968). It is unclear whether there were other characteristics of the children treated before 10 months of age (later birth cohort?) that might help to account for this finding.

Findings from these early, pre-1965 cohorts are heterogeneous. Two Canadian studies with a total of 73 subjects (Hanley et al., 1971; Scriver et al., 1968) reported a mean IQ of approximately 55, as did a small German study with 3 late-treated subjects (Schweitzer et al., 1987). In contrast, four studies with data on a total of 124 late-treated subjects from the United States, England, and Germany reported a pooled mean IQ of 70.2 (Dobson et al., 1968; Legido et al., 1993; Siegel et al., 1968; Smith et al., 1978).

Studies that reported cognitive assessments among children with PKU who were born in high-income countries after the mid-1960s are summarized in Table 1. Campistol Plana, Alvarez Dominguez, Riverola de Veciana, Castillo Rivera, and Giner Soria (1991) reported that of 7 children 4–14 years of age with late-treated PKU who were born in Catalonia, Spain, 1976–1985, and diagnosed at 1.5–8 years of age, 5 who were on a low-Phe diet had a mean IQ score of 63 compared with a mean IQ of 48 for 2 children who were untreated. A subsequent report from the same Barcelona hospital presented information on a subset of 11 patients with tremor among 18 adult patients (19–37 years of age) with late-treated PKU. Of those 11, 8 had been placed on a low-Phe diet at 0.3–6.75 years of age and had a mean IQ score of 69 (range 50–103; Pérez-Dueñas et al., 2005). Seven patients without tremor, whose individual scores were not listed, had IQ scores higher by 10 points on average.

A study from Andalucía, Spain, by Lluch Fernández et al. (1988) reported data on 9 children with late-treated PKU who were diagnosed at 0.5–8 years of age and born 1968–1980. When assessed in 1985 at 5–18 years of age, 7 (78%) had IQ scores ≤ 70, and the other 2 had scores above 80. However, one third of the children were on uncontrolled diets, and scores were not reported separately for the children with low-Phe diets. All 9 children were identified with PKU after being diagnosed with developmental delays, which raises the problem of ascertainment bias.

In Italy, investigators analyzed cognitive data on 11 late-treated subjects with PKU who were diagnosed during 1965–1992 at 0.8–4.5 years of age and remained on a low-Phe diet for at least 4 years, in addition to a 2-year-old child who had only recently been put on a low-Phe diet (Leuzzi, Trasimeni, Gualdi, & Antonozzi, 1995). These subjects were recruited to participate in an MRI study and were not necessarily representative. Three (27%) of the 11 late-treated subjects who had been treated for at least 4 years had IQ > 85, 4 (36%) had IQ 56–65, and 4 (36%) had IQ ≤ 50. Although the

Table 1
Mean IQ Scores and Percentages ≤ 70 and ≤ 50 for Late-Treated Individuals With PKU Following a Low-Phe Diet for Several Years

Study	Country	Birth cohort(s)	Age at Dx (years)	Age at testing (years)	*N*	Mean IQ	% IQ < 70	% IQ ≤ 50
Lluch Fernández et al. (1988)	Spain	1968–1980	0.5–8	5–18	9	NR	78	NR
Campistol Plana et al. (1991)	Spain	1976–1985	1.5–8	4–14	5	63	100	20
Pérez-Dueñas et al. (2005)	Spain	NR	0.3–6.8	19–36	8	69	62	12
Leuzzi et al. (1995)	Italy	NR	0.8–4.5	6.8–25	11	65	73	36
	Italy	NR	0.8–4.5	17.5–25	7	51	100	57
	Italy	NR	1.5–2.4	6.8–15	4	86	25	0

Note. IQ = intelligence quotient; NR = not reported; low-Phe diet = low-phenylalanine diet; PKU = phenylketonuria.

authors did not report this, there was an indication of a likely cohort effect. The 7 late-treated individuals who were ages 17.5–25 years at the time of the final assessment all had IQ < 70, many of whom were born prior to the introduction of newborn screening. In comparison, only 1 (25%) of 4 children, aged 6.8–15 years who had been treated for multiple years had IQ < 70, and their average IQ score was 86. All of the late-treated individuals had initially been diagnosed with cognitive deficits, which reflects ascertainment bias. Several of those children subsequently had cognitive tests in the normal range, as noted by the authors.

A large U.S. study with data for 49 late-treated adults in California who were followed to age 18 years (Koch et al., 1999) provides data that allow for an analysis of differences in IQ scores across birth cohorts (Table 2). Overall, 27% of those with IQ scores measured had an IQ ≤ 70, 50% among those born in 1960 or earlier and 14%–16% among those born after 1960. Among the post-1965 birth cohort, IQ scores below 70 were found only among those diagnosed at 3 years of age or older. Mean adult IQ scores showed a similar pattern, 72.9 for the 1951–1960 cohort and 84.1–84.3 for the post-1960 cohorts. Among the post-1965 cohort, a mean IQ of 76.2 was observed for 5 subjects diagnosed at 3–7 years of age compared with a mean IQ of 92.0 for 5 subjects diagnosed at 1–2 years of age. Individuals who began treatment after 3 months but before 12 months of age had a mean adult IQ of 91.8 in the 1961–1965 birth cohort and 97.3 in the 1965–1978 birth cohort. The IQ scores reported in this study are substantially higher than those for contemporary cohorts reported in Japanese and Canadian studies of adults with late-treated PKU who had been treated as young children (Izumi et al., 2006; Netley et al., 1984). This might reflect differences in the strictness of dietary control achieved across populations and centers. The California findings indicate cognitive outcomes in late-treated PKU that are achievable with rigorous dietary control.

Table 2

Mean IQ Scores and Percentages < 70 Among 49 Adults With Classic PKU Born 1951–1978, Diagnosed After 3 Months of Age, and Followed Up at the University of California at Los Angeles (Koch et al., 1999)

		Age at diagnosis			
Birth year		< 1 year	1–2 years	3 years+	Any age
1951–1960	Mean	69.0	78.4	59.0	72.9
	IQ < 70 (%)	100	36	75	50
	Proportion	1/1	4/11	3/4	8/16
1961–1965	Mean	91.8	79.2	75.8	84.3
	IQ < 70 (%)	0	20	40	16
	Proportion	0/9	1/5	2/5	3/19
1966–1978	Mean	97.3	92.0	76.2	84.1
	IQ < 70 (%)	0	0	40	14
	Proportion	0/4	0/5	2/5	2/14
All years	Mean				
	IQ < 70 (%)	7	24	50	27
	Proportion	1/14	5/21	7/14	13/49

Note. IQ = intelligence quotient; PKU = phenylketonuria.

Longitudinal assessments. The methodological challenges in interpreting longitudinal assessments differ from cross-sectional assessments. Ascertainment bias is less of an issue. Only children with an initial developmental delay contribute data relevant to the question of partial reversibility of delays. The quality of dietary control and its duration are important. Short- and medium-term assessments provide limited information because it might take several years on a low-Phe diet for improvements to fully manifest themselves.

Problems of comparability of cognitive assessments are a particular challenge for longitudinal assessments. Studies that report changes in mean test scores for different groups based on treatment and age of initiation of a low-Phe diet are more informative because the between-group comparisons allow one to draw inferences unaffected by test type or the Flynn effect. Because valid IQ scores cannot be obtained from very young children, assessments of changes over time involve the comparison of DQ and IQ scores. Only DQ scores assessed after 12 months of age are included in this section because scores under 12 months do not measure mental development (Slater, 1997). Waisbren et al. (2000) compared DQ scores measured at 2 years of age with the Bayley Scales Mental Index to those assessed at 4 years of age with the McCarthy Scales General Cognitive Index for children of mothers with HPA and a control group and reported that mean DQ decreased from 96 to 85 for the former and from 114 to 110 for the latter. Waisbren and Azen (2003) reported IQ scores for the same cohort measured at 7 years of age; mean IQ scores were 90 for the HPA offspring and 106 for the control group. Despite differences in test types, neither group showed significant changes over time in cognitive test scores.

Older studies often reported appreciable increases in cognitive test scores among late-diagnosed children with PKU. Fuller and Shuman (1969) commented that some children put on a low-Phe diet at 3 years of age improved sufficiently to outperform other children who had been placed on a low-Phe diet soon after birth. Knox (1972) summarized data on changes in either DQ or IQ after at least 3 years of treatment for five groups of children for whom the mean age of diagnosis was 10–30 months, with mean changes ranging from barely imperceptible to over 25 points. McBean and Stephenson (1968) reported longitudinal assessments in 3 children in Glasgow, Scotland, treated beginning at 35–42 months of age. The increases in IQ scores ranged from 25 to 41 points (from 55 to 96, from 25 to 62, and from 52 to 77). One other child treated at 67 months of age showed no change in IQ scores.

Berman et al. (1966) reported repeated assessments for 22 treated and 5 untreated children with PKU seen at a Wisconsin clinic 1957–1964. The mean change in test scores was +7.5 points for the treated group, compared with a change of −8 points for those who were untreated. For 15 children who were at least 6 months of age at the initial assessment and with a final assessment more than 12 months after the initial assessment, the mean change was +11.6 points, with a range from −3 to +35 points.

A study in East Germany involving 19 children who were diagnosed with PKU after 6 months of age and followed for an average of 5 years while on a low-Phe diet reported that DQ scores for 7 children who were more than 24 months of age when first treated increased by an average of 13.4 points before the children went off the diet ($p < .05$; Theile, Bührdel, & Graustein, 1976). No significant increase in DQ scores occurred among 12 children put on the diet before age 24 months, but these children had a lesser degree of impairment. Only 5 of 12 had initial DQ ≤ 30 compared with 5 of 7 children in the older group.

A long-term study of changes in cognitive test scores conducted at the University of California at Los Angeles reported cognitive assessments at both age of diagnosis and during adulthood for 30 individuals born 1951–1978 and diagnosed at 12 months of age or later who followed a low-Phe diet (Koch et al., 1999). Changes in test score for the 29 subjects for whom test scores at time of diagnosis were available ranged from 29 to 70. The mean change in test scores between diagnosis and adult assessment was 24.4 (range = 2–48 points) for 12 subjects born 1951–1960, 20.8 (range = 4–34 points) for 8 subjects born 1961–1965, and 36.0 (range = 15–56 points) for 10 subjects born 1965–1978. In each birth cohort, there was no significant difference in mean increments in test scores between those diagnosed at 1–2 years of age and those diagnosed after 2 years of age.

In Mexico City, investigators followed 5 children with PKU who began treatment after approximately 12 months of age (Carnevale, Velazquez, Ruiz, & Del Castillo, 1979). The investigators used the Gesell and Terman-Merrill to measure DQ or IQ; the results were presented graphically. Four of the children followed a low-Phe diet for at least 12 months. The increments were approximately 30, 25, 5, and 30 points; the last observation was for a child who had a DQ of about 35 when diagnosed at 3.5 years of age and had an IQ of about 65 when retested at 5 years of age. Conversely, a child who never established dietary control continued to have a DQ under 20.

In Hamburg, Germany, investigators followed 8 late-treated children who were put on treatment after 6 months of age and conducted repeated assessments using the Columbia Mental Maturity Scale (Steinhausen & Börner, 1977). Only raw results were reported and they were presented graphically. The mean score increased between the first and third measurements by about 6 points, from 27 to 33, which on a scale of 100 would be approximately twice as large. Also, the majority of the late-treated children had poorly controlled diets. Longitudinal results were not reported separately for the children who experienced good dietary control, but their test scores were much better, about 45 on average.

Yang et al. (1999) reported on the experience of 161 late-diagnosed children seen at a hospital in Beijing, China, who were on a low-Phe diet for anywhere from 3 months to 6 years. Most (73.3%) showed marked improvement, including 21.1% (36 of 161) who had a final IQ score of 75 or greater. The authors noted that the low-Phe diet available at the time was suboptimal. Indeed, children with PKU in that center whose treatment was initiated before age 3 months were no more likely to have IQ > 75 (Yang et al., 1999).

Two publications from another Beijing hospital also contained longitudinal follow-up data. First, Yu et al. (2003) reported that 81 children with late-treated PKU were maintained on a low-Phe diet for 10–18 months. Their mean Gesell IQ score was reported to have increased from 61 to 72 ($p < .05$). A subsequent publication from the same institution reported detailed assessments of a small group of 19 children with late-treated PKU who at ages 1 through 6 years were given a low-Phe diet and followed for an average of 9 months (Wang, Zhou, & Yu, 2006). These children were assessed before and after with both Gesell's intelligence scale and MRI. The following calculations are based on Table 1 of Wang et al. (2006). The mean cognitive test score increased by 15.8 points, from an initial mean of 44.6 to a final mean of 60.4 ($p < .01$). Individual children had gains ranging from 6 to 32 points. Most showed WM lesions that underwent significant decline following dietary treatment, although this could have been coincidental.

Other investigators have reported longitudinal assessments for case reports. For example, Holmgren, Blomquist, and Samuelson (1980) reported that a Swedish child with PKU was put on a low-Phe diet for 8 years beginning at age 8 years and the child's IQ score rose by 41 points, from 59 to 90. Investigators in Poland, where screening for PKU began in the 1980s, reported long-term assessments on 2 late-diagnosed siblings with the same *PAH* genotype (Gizewska et al., 2003). The older sibling was diagnosed at 4.5 years of age, with an IQ of 63. After 14 years of dietary treatment, this child's IQ had increased by 44 points, to 107. The younger sibling, diagnosed and put on treatment at 11 months of age, was unable to be tested with an IQ test at 4 years of age because of severe impairment, with no subsequent improvement through 22 years of age. The 2 siblings had similar blood Phe levels, but the brain Phe levels differed (Gizewska et al., 2003).

Discussion

The conventional wisdom that PKU uniformly results in severe intellectual disability in the absence of newborn screening properly should refer to outcomes in the absence of treatment during childhood. Children who are put on a low-Phe diet when diagnosed often reach adulthood with lesser degrees of impairment. Older studies that reported mean IQ scores in the 50s among individuals with late-treated PKU were subject to ascertainment bias (Birch & Tizard, 1967); other studies from high-income countries reported mean IQ scores of approximately 70 (Dobson et al., 1968; Legido et al., 1993; Siegel et al., 1968; Smith et al., 1978). Most children missed by universal newborn screening programs in high-income countries in recent decades and described in the scientific literature have had IQ scores in the borderline to the low-normal range (Campistol Plana et al., 1991; Koch et al., 1999; Leuzzi et al., 1995; Trefz et al., 2000).

The level of cognitive test scores for late-treated individuals with PKU born in low- or middle-income countries is often much lower than for those born in high-income countries who were missed by newborn screening. Thus, for example, 5 late-treated children with PKU in Turkey were all reported to have IQ < 50 (Ozalp et al., 1994). Similarly unfavorable results were reported from East German (Wässer, Theile, Graustein, & Bührdel, 1979) and Romania (Popescu et al., 2000). Likewise, the absolute levels of test scores in the longitudinal assessments from East Germany (Theile et al., 1976), Mexico (Carnevale et al., 1979), and China (Wang et al., 2006; Yang et al., 1999; Yu et al., 2003) were relatively low, despite favorable changes during the course of treatment.

Longitudinal data on cognitive test scores provide evidence of partial reversibility of the cognitive impairment associated with late-diagnosed PKU. Studies that have conducted repeated developmental assessments of late-treated PKU subjects have reported marked improvement in standardized test scores in many cases, although this is highly variable among individuals. This is not a new finding but reinforces observations first reported by Berman et al. (1966), "While phenylketonuric children showed significant gains in developmental rate and IQ during dietary treatment, the fact that they did not attain rates as high as their unaffected siblings suggests that some of the effects of PKU continue to be operative even when blood Phe levels are controlled." More recent studies have demonstrated that some individuals achieve gains in test scores of 20–45 points, although others manifest no improvement despite continued treatment (Gizewska et al., 2003). In comparison, longitudinal studies of early-treated PKU subjects, as well as unaffected individuals, have reported that test scores were stable or, when dietary compliance became more relaxed, in some cases declined (Burgard, Schmidt, Rupp, Schneider, & Bremer, 1996). None of the studies reviewed here discussed the potential contribution of behavioral and other nondietary interventions to the achievement of improved cognitive test scores among many individuals with late-treated PKU.

The pattern of longitudinal increases in standardized test scores among children with late-treated PKU appears to resemble that of children born extremely low birth weight, a finding, which has been suggested to reflect the plasticity of brain development in early to mid-childhood (Hack et al., 2005). The latter study found that among children with developmental delay at 20 months of age, 20 (44%) of 45 with intact neurosensory function had IQ scores ≥ 85 at 8 years of age, as did 8 (24%) of 33 with neurosensory impairments. Long-term follow-up of children with low DQ scores is needed before a definitive assessment can be made. A DQ score of 50 for a young child does not mean that the child will necessarily have intellectual disability. On the other hand, although there may be partial reversibility of cognitive deficits, late-treated PKU is associated with irreversible neurological damage that seriously reduces individual abilities.

The improvement in cognitive test scores often observed with continued treatment of late-diagnosed children with PKU should give some degree of hope to parents of children who unfortunately missed being detected through newborn screening (Levy, 2000). It in no way reduces the urgency of universal screening and prompt initiation of treatment, which is essential for optimal development. A delay of 4 weeks in starting treatment has been associated with a loss of several IQ points at age 4 years (Smith, Beasley, & Ades, 1990). Reduced IQ scores represent an unnecessary and avoidable loss of cognitive and economic potential. Other aspects of neurodevelopment, such as visual–spatial executive functions, are often compromised among individuals with PKU whose blood Phe levels were not strictly controlled, even those with IQs in the normal range (Burgard, 2000; Diamond & Herzberg, 1996; Waisbren, Brown, de Sonneville, & Levy, 1994). Early diagnosis in combination with intensive treatment is the only means to minimize neurological damage caused by PKU. Promising alternative treatments to dietary Phe restriction include large neutral amino acid administration and supplementation with the PAH cofactor tetrahydrobiopterin (BH_4; Scriver, 2007).

References

Abadie, V., Berthelot, J., Feillet, F., Maurin, N., Mercier, A., de Baulny, H. O., et al. (2001). Neonatal screening and long-term follow-up of phenylketonuria: The French database. *Early Human Development*, *65*, 149–158.

Anderson, P. J., Wood, S. J., Francis, D. E., Coleman, L., Anderson, V., & Boneh, A. (2007). Are neuropsychological impairments in children with early-treated phenylketonuria (PKU) related to white matter abnormalities or elevated phenylalanine levels? *Developmental Neuropsychology*, *32*, 645–668.

Anderson, P. J., Wood, S. J., Francis, D. E., Coleman, L., Warwick, L., Casanelia, S., et al. (2004). Neuropsychological functioning in children with early-treated phenylketonuria: Impact of white matter abnormalities. *Developmental Medicine and Child Neurology*, *46*, 230–238.

Berman, P. W., Waisman, H. A., & Graham, F. K. (1966). Intelligence in treated phenylketonuric children: A developmental study. *Child Development*, *37*, 731–747.

Bhasin, T. K., Brocksen, S., Avchen, R. N., & Van Naarden Braun, K. (2006). Prevalence of four developmental disabilities among children aged 8 years—Metropolitan Atlanta Developmental Disabilities Surveillance Program, 1996 and 2000. *Morbidity and Mortality Weekly Report Surveillance Summaries*, *55*, 1–9.

Birch, H. G., & Tizard, J. (1967). The dietary treatment of phenylketonuria: Not proven? *Developmental Medicine and Child Neurology*, *9*, 9–12.

Bower, C., Leonard, H., & Petterson, B. (2000). Intellectual disability in Western Australia. *Journal of Paediatrics and Child Health*, *36*, 213–215.

Brumm, V. L., Azen, C., Moats, R. A., Stern, A. M., Broomand, C., Nelson, M. D., et al. (2004). Neuropsychological outcome of subjects participating in the PKU adult collaborative study: A preliminary review. *Journal of Inherited Metabolic Disease*, *27*, 549–566.

Burgard, P. (2000). Development of intelligence in early treated phenylketonuria. *European Journal of Pediatrics*, *159*, S74–S79.

Burgard, P., Schmidt, E., Rupp, A., Schneider, W., & Bremer, H. J. (1996). Intellectual development of the patients of the German Collaborative Study of children treated for phenylketonuria. *European Journal of Pediatrics*, *155*, S33–S38.

Campistol Plana, J., Alvarez Dominguez, L., Riverola de Veciana, A. T., Castillo Rivera, P., & Giner Soria, P. (1991). [Hyperphenylanalinemia and phenylketonuria: The importance of early diagnosis and follow up at a health center]. *Anales Españoles de Pediatría*, *34*, 51–56.

Carnevale, A., Velazquez, A., Ruiz, F., & Del Castillo, V. (1979). [Management of patients with phenylketonuria in Mexico]. *Boletín Médico del Hospital Infantil de México*, *36*, 375–384.

Cleary, M. A., Walter, J. H., Wraith, J. E., White, F., Tyler, K., & Jenkins, J. P. (1995). Magnetic resonance imaging in phenylketonuria: Reversal of cerebral white matter change. *Journal of Pediatrics*, *127*, 251–255.

Diamond, A. (1996). Evidence for the importance of dopamine for prefrontal cortex functions early in life. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *351*, 1483–1493.

Diamond, A., Ciaramitaro, V., Donner, E., Djali, S., & Robinson, M. (1994). An animal model of early-treated PKU. *Journal of Neuroscience*, *14*, 3072–3082.

Diamond, A., & Herzberg, C. (1996). Impaired sensitivity to visual contrast in children treated early and continuously for phenylketonuria. *Brain*, *119*, 523–538.

Diamond, A., Prevor, M. B., Callender, G., & Druin, D. P. (1997). Prefrontal cortex cognitive deficits in children treated early and continuously for PKU. *Monographs of the Society for Research in Child Development*, *62*.

Dobson, J., Koch, R., Williamson, M., Spector, R., Frankenburg, W., O'Flynn, M., et al. (1968). Cognitive development and dietary therapy in phenylketonuric children. *New England Journal of Medicine, 278*, 1142–1144.

Feldman, W. (1994). Screening for phenylketonuria. In Canadian Task Force on the Periodic Health Examination (Ed.), *Canadian guide to clinical preventive health care* (pp. 180–188). Ottawa: Health Canada.

Flynn, J. R. (1999). Searching for justice: The discovery of IQ gains over time. *American Psychologist, 54*, 5–20.

Fuller, R. N., & Shuman, J. B. (1969). Phenylketonuria and intelligence: Trimodal response to dietary treatment. *Nature, 221*, 639–642.

Gizewska, M., Cabalska, B., Cyrytowski, L., Nowacki, P., Zekanowski, C., Walczak, M., et al. (2003). Different presentations of late-detected phenylketonuria in two brothers with the same R408W/R111X genotype in the PAH gene. *Journal of Intellectual Disability Research, 47*, 146–152.

Hack, M., Taylor, H. G., Drotar, D., Schluchter, M., Cartar, L., Wilson-Costello, D., et al. (2005). Poor predictive validity of the Bayley Scales of Infant Development for cognitive function of extremely low birth weight children at school age. *Pediatrics, 116*, 333–341.

Hanley, W. B. (2007). Phenylketonuria: Questioning the gospel. *Expert Review of Endocrinology and Metabolism, 2*, 809–816.

Hanley, W. B., Linsao, L. S., & Netley, C. (1971). The efficacy of dietary therapy for phenylketonuria. *Canadian Medical Association Journal, 104*, 1089–1091.

Holmgren, G., Blomquist, H. K., & Samuelson, G. (1980). Positive effect of a late introduced modified diet in an 8-year-old PKU child. *Neuropadiatrie, 11*, 76–79.

Izumi, M., Yamazaki, H., Nakabayashi, H., & Owada, M. (2006). [Magnetic resonance imaging of the brain in phenylketonuria]. *No To Hattatsu, 38*, 27–31.

Kaye, C. I., & the Committee on Genetics, American Academy of Pediatrics. (2006). Newborn screening fact sheets. *Pediatrics, 118*, e934–e963.

Knox, W. E. (1972). Phenylketonuria. In J. B. Stanbury, J. B. Wyngaarde, & D. S. Frederickson (Eds.), *The metabolic basis of inherited disease* (pp. 266–295). New York: McGraw-Hill.

Koch, R., Burton, B., Hoganson, G., Peterson, R., Rhead, W., Rouse, B., et al. (2002). Phenylketonuria in adulthood: A collaborative study. *Journal of Inherited Metabolic Disease, 25*, 333–346.

Koch, R., Moats, R., Guttler, F., Guldberg, P., & Nelson, M., Jr. (2000). Blood–brain phenylalanine relationships in persons with phenylketonuria. *Pediatrics, 106*, 1093–1096.

Koch, R., Moseley, K., Ning, J., Romstad, A., Guldberg, P., & Guttler, F. (1999). Long-term beneficial effects of the phenylalanine-restricted diet in late-diagnosed individuals with phenylketonuria. *Molecular Genetics and Metabolism, 67*, 148–155.

Legido, A., Tonyes, L., Carter, D., Schoemaker, A., Di George, A., & Grover, W. D. (1993). Treatment variables and intellectual outcome in children with classic phenylketonuria. A single-center-based study. *Clinical Pediatrics, 32*, 417–425.

Leuzzi, V., Tosetti, M., Montanaro, D., Carducci, C., Artiola, C., Carducci, C., et al. (2007). The pathogenesis of the white matter abnormalities in phenylketonuria. A multimodal 3.0 tesla MRI and magnetic resonance spectroscopy (1H MRS) study. *Journal of Inherited Metabolic Disease, 30*, 209–216.

Leuzzi, V., Trasimeni, G., Gualdi, G. F., & Antonozzi, I. (1995). Biochemical, clinical and neuroradiological (MRI) correlations in late-detected PKU patients. *Journal of Inherited Metabolic Disease, 18*, 624–634.

Levy, H. L. (2000). Comments on final intelligence in late treated patients with phenylketonuria. *European Journal of Pediatrics, 159*, S149.

Levy, H. L., & Albers, S. (2000). Genetic screening of newborns. *Annual Review of Genomics and Human Genetics, 1*, 139–177.

Lluch Fernández, M. D., Ramos Sánchez, I., Marchante Cobos, C., Pérez Pérez, G., Martínez Martínez, J. J., & Estefanía Gallardo, C. (1988). [Phenylketonuria. Treatment and developmental control]. *Anales Españoles de Pediatría, 28*, 327–330.

McBean, M. S., & Stephenson, J. B. (1968). Treatment of classical phenylketonuria. *Archives of Disease in Childhood, 43*, 1–7.

Möller, H. E., Weglage, J., Bick, U., Wiedermann, D., Feldmann, R., & Ullrich, K. (2003). Brain imaging and proton magnetic resonance spectroscopy in subjects with phenylketonuria. *Pediatrics, 112*, 1580–1583.

Möller, H. E., Weglage, J., Wiedermann, D., & Ullrich, K. (1998). Blood-brain barrier phenylalanine transport and individual vulnerability in phenylketonuria. *Journal of Cerebral Blood Flow and Metabolism, 18*, 1184–1191.

Netley, C., Hanley, W. B., & Rudner, H. L. (1984). Phenylketonuria and its variants: Observations on intellectual functioning. *Canadian Medical Association Journal, 131*, 751–755.

Ozalp, I., Coşkun, T., Ozgüç, M., Tokatli, A., Yalaz, K., Vanli, L., et al. (1994). Genetic and neurological evaluation of untreated and late-treated patients with phenylketonuria. *Journal of Inherited Metabolic Disorders, 17*, 371.

Pérez-Dueñas, B., Pujol, J., Soriano-Mas, C., Ortiz, H., Artuch, R., Vilaseca, M. A., et al. (2006). Global and regional volume changes in the brains of subjects with phenylketonuria. *Neurology, 66*, 1074–1078.

Pérez-Dueñas, B., Valls-Solé, J., Fernández-Alvarez, E., Conill, J., Vilaseca, M. A., Artuch, R., et al. (2005). Characterization of tremor in phenylketonuric subjects. *Journal of Neurology, 252*, 1328–1334.

Pietz, J., Kreis, R., Schmidt, H., Meyding-Lamade, U., Rupp, A., & Boesch, C. (1996). Phenylketonuria. Findings at MR imaging and localized in vivo H-1 MR spectroscopy of the brain in patients with early treatment. *Radiology, 201*, 413–420.

Popescu, T., Popescu, A., Oltean, L., Kozak, L., Blazkova, M., & Jebeleanu, G. (2000). Retrospective neuropsychological evaluation of 18 Romanian subjects with hyperphenylalaninaemia. *Acta Paediatrica, 89*, 1502–1503.

Ramus, S. J., Forrest, S. M., Pitt, D. D., & Cotton, R. G. (1999). Genotype and intellectual phenotype in untreated phenylketonuria patients. *Pediatric Research, 45*, 474–481.

Roeleveld, N., Zielhuis, G. A., & Gabreels, F. (1997). The prevalence of mental retardation: A critical review of recent literature. *Developmental Medicine and Child Neurology, 39*, 125–132.

Schweitzer, S., Wolff, G., & Brodehl, J. (1987). [Psychosocial aspects of the treatment of phenylketonuria]. *Klinische Pädiatrie, 199*, 90–94.

Scriver, C. R. (2007). The *PAH* gene, phenylketonuria, and a paradigm shift. *Human Mutation, 28*, 831–845.

Scriver, C. R., Katz, L., & Clow, C. (1968). Phenylketonuria and diet. *Canadian Medical Association Journal, 98*, 124–125.

Scriver, C. R., & Kaufman, S. (2001). Hyperphenylalanemia: Phenylalanine hydroxylase deficiency. In C. R. Scriver, A. L. Beaudet, W. S. Sly, & D. Valle (Eds.), *Bases of inherited disease* (pp. 1667–1724). New York: McGraw-Hill.

Seymour, C. A., Thomason, M. J., Chalmers, R. A., Addison, G. M., Bain, M. D., Cockburn, F., et al. (1997). Newborn screening for inborn errors of metabolism: A systematic review. *Health Technology Assessment, 1*(11), 1–95.

Siegel, F. S., Balow, B., Fisch, R. O., & Anderson, V. E. (1968). School behavior profile ratings of phenylketonuric children. *American Journal of Mental Deficiency, 72*, 937–943.

Slater, A. (1997). Can measures of infant habituation predict later intellectual ability? *Archives of Disease in Childhood, 77*, 474–476.

Smith, I., Beasley, M. G., & Ades, A. E. (1990). Intelligence and quality of dietary treatment in phenylketonuria. *Archives of Disease in Childhood, 65*, 472–478.

Smith, I., Lobascher, M. E., Stevenson, J. E., Wolff, O. H., Schmidt, H., Grubel-Kaiser, S., et al. (1978). Effect of stopping low-phenylalanine diet on intellectual progress of children with phenylketonuria. *British Medical Journal, 2*, 723–726.

Smith, I., & Wolff, O. H. (1974). Natural history of phenylketonuria and influence of early treatment. *Lancet, 2*, 540–544.

Steinhausen, H. C., & Börner, S. (1977). [Treatment of phenylketonuria: Psychological results of a 5-year follow-up study]. *Monatsschrift für Kinderheilkunde, 125*, 190–194.

Stemerdink, B. A., Kalverboer, A. F., van der Meere, J. J., van der Molen, M. W., Huisman, J., de Jong, L. W., et al. (2000). Behaviour and school achievement in patients with early and continuously treated phenylketonuria. *Journal of Inherited Metabolic Disorders, 23*, 548–562.

Theile, H., Bührdel, P., & Graustein, I. (1976). [Development of children with late treatment of phenylketonuria following diet therapy]. *Kinderärztliche Praxis, 44*, 338–343.

Thompson, A. J., Tillotson, S., Smith, I., Kendall, B., Moore, S. G., & Brenton, D. P. (1993). Brain MRI changes in phenylketonuria Associations with dietary status. *Brain, 116*, 811–821.

Trefz, F. K., Cipcic-Schmidt, S., & Koch, R. (2000). Final intelligence in late treated patients with phenylketonuria. *European Journal of Pediatrics, 159*, S145–S148.

Ullrich, K., Möller, H., Weglage, J., Schuierer, G., Bick, U., Ludolph, A., et al. (1994). White matter abnormalities in phenylketonuria: Results of magnetic resonance measurements. *Acta Paediatrica Supplement, 407*, 78–82.

Van Naarden Braun, K., Yeargin-Allsopp, M., Schendel, D., & Fernhoff, P. (2003). Long-term developmental outcomes of children identified through a newborn screening program with a metabolic or endocrine disorder: A population-based approach. *Journal of Pediatrics, 143*, 236–242.

Waisbren, S. E., & Azen, C. (2003). Cognitive and behavioral development in maternal phenylketonuria offspring. *Pediatrics, 112*, 1544–1547.

Waisbren, S. E., Brown, M. J., de Sonneville, L. M., & Levy, H. L. (1994). Review of neuropsychological functioning in treated phenylketonuria: An information processing approach. *Acta Paediatrica Supplement, 407*, 98–103.

Waisbren, S. E., Hanley, W., Levy, H. L., Shifrin, H., Allred, E., Azen, C., et al. (2000). Outcome at age 4 years in offspring of women with maternal phenylketonuria: The Maternal PKU Collaborative Study. *JAMA, 283*, 756–762.

Waisbren, S. E., Noel, K., Fahrbach, K., Cella, C., Frame, D., Dorenbaum, A., et al. (2007). Phenylalanine blood levels and clinical outcomes in phenylketonuria: A systematic literature review and meta-analysis. *Molecular Genetics and Metabolism, 92*, 63–70.

Waisbren, S. E., Schnell, R. R., & Levy, H. L. (1980). Diet termination in children with phenylketonuria: A review of psychological assessments used to determine outcome. *Journal of Inherited Metabolic Disease, 3*, 149–153.

Wang, Z. X., Zhou, Z. S., & Yu, W. M. (2006). [Brain white matter lesions of children with phenylketonuria before and after treatment]. *Zhongguo Dang Dai Er Ke Za Zhi, 8*, 13–16.

Wässer, S., Theile, H., Graustein, I., & Bührdel, P. (1979). [Significance of early detection and early treatment of the development of children with phenylketonuria]. *Beihefte Psychiatrie, Neurologie und medizinische Psychologie, 25*, 137–149.

Weglage, J., Wiedermann, D., Denecke, J., Feldmann, R., Koch, H. G., Ullrich, K., et al. (2002). Individual blood-brain barrier phenylalanine transport in siblings with classical phenylketonuria. *Journal of Inherited Metabolic Disease*, *25*, 431–436.

Yang, Y., Gu, Q., Zhang, Z., Mi, C., Wang, L., & Wu, X. (1999). A clinical investigation of 228 patients with phenylketonuria in mainland China. *Southeast Asian Journal of Tropical Medicine and Public Health, 30,* 58–60.

Yu, W. M., Xu, L., Li, X. W., He, C., Shen, M., Zhang, Z. X., et al. (2003). [An eighteen-year study on phenylketonuria]. *Zhongguo Yi Xue Ke Xue Yuan Xue Bao, 25,* 218–222.

Child Development, January/February 2010, Volume 81, Number 1, Pages 212–223

Placement in Foster Care Enhances Quality of Attachment Among Young Institutionalized Children

Anna T. Smyke and
Charles H. Zeanah
Tulane University School of Medicine

Nathan A. Fox
University of Maryland

Charles A. Nelson
Harvard Medical School/Children's Hospital Boston

Donald Guthrie
University of California Los Angeles

This study examined classifications of attachment in 42-month-old Romanian children ($N = 169$). Institutionalized since birth, children were assessed comprehensively, randomly assigned to care as usual (CAU) or to foster care, and compared to family-reared children. Attachment classifications for children in foster care were markedly different from those in the CAU. Importantly, children placed in foster care before 24 months were more likely to have secure attachments and if placed earlier were less likely to have disorganized or insecure-other attachments. Cognitive status predicted greater likelihood of organized attachment in the CAU and greater likelihood of secure attachment in the foster care and never-institutionalized groups. Foster care is an important intervention to reduce the adverse effects following early deprivation.

Attachment theory implies that the formation of a preferred attachment between infant and parent reflects the activity of ''experience expectant'' neural systems (Greenough, Black, & Wallace, 1987). Specifically, our species has evolved in such a way as to expect that infants will be cared for by a capable caregiver, one who can ensure survival. Accordingly, only in the most extreme conditions of rearing do attachments fail to develop, such as in children who have been severely neglected (Boris et al., 2004; Zeanah et al., 2004) or in those raised in institutional settings (Smyke, Dumitrescu, & Zeanah, 2002; Tizard & Rees, 1975; Zeanah, Smyke, Koga, Carlson, & the BEIP Core Group, 2005). In both instances, the young child's ability to form selective attachments is limited by aberrant environmental conditions. Even in these atypical settings, however, attachment to a preferred caregiver seems often to occur, indicating a ''hard-wired'' propensity for the young child to become attached.

On the other hand, the type or quality of attachments a young child forms seems to reflect neural processes that are ''experience dependent'' (Greenough et al., 1987). That is, attachment theory posits that the young child constructs attachment relationships that reflect the child's experiences with particular caregivers. Quality of attachment is typically assessed by using the Strange Situation procedure (SSP; Ainsworth, Blehar, Waters, & Wall, 1978) that examines individual differences in the organization of attachment and exploratory behaviors. As a child forms an attachment to an adult caregiver, experiences with that adult seem to relate to the type or quality of that attachment. This model is supported by various lines of evidence about quality of attachment.

In addition to the fact that children may develop different types of attachments to different caregivers, a large number of studies have linked parental characteristics to infant attachment, including parents' sensitive behavior toward the infant, state of mind with regard to attachment, insightfulness, and reflective functioning (De Wolff & van IJzendoorn,

---

The Bucharest Early Intervention Project (BEIP) was funded by the John D. and Catherine T. MacArthur Foundation Research Network on Early Experience and Brain Development (Charles A. Nelson, network chair). We thank our dedicated BEIP research staff as well as Hermi R. Woodward and Sebastian F. Koga. We are very grateful to the children and caregivers who have participated in the BEIP.

Correspondence concerning this article should be addressed to Anna T. Smyke, Department of Psychiatry/Neurology, Tulane University Health Sciences Center, 1440 Canal Street – TB-52, New Orleans, LA 70112. Electronic mail may be sent to asmyke@tulane.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8101-0014

1997; Koren-Karie, Oppenheim, & Doley, 2002; Slade, Sadler, & Mayes, 2005; van IJzendoorn, 1995).

Nevertheless, the strongest evidence to date demonstrating that the specific quality of the child's attachment depends upon parental characteristics comes from various interventions in which changing parents' behaviors or their states of mind have led to changes in the quality of attachment of the child to that caregiver. These interventions have included impoverished mothers and infants (Heinicke, Fineman, & Ruth, 1999), mothers and their temperamentally difficult (van den Boom, 1994) or highly reactive (Klein Velderman, Bakermans-Kranenburg, & Juffer, 2006) infants, maltreated infants and their mothers (Cicchetti, Rogosch, & Toth, 2006; Toth, Maughan, & Manly, 2002), and foster mothers and their infants (Dozier, Grasso, & Lindhiem, 2007; Fisher & Kim, 2007). Impressively, these studies have shown uniformly better quality of attachments between infants and caregivers in intervention rather than endangered or stressed control groups.

There is also evidence about the ability of infants and toddler raised in institutional settings to recover from their early adversity and form healthy attachments. Some of the young children being raised in institutions have no attachments to their institutional caregivers (Smyke et al., 2002; Tizard & Rees, 1975; Zeanah et al., 2005), and a majority have compromised attachments, as reflected by the prevalence of disorganized and other atypical classifications of attachment (Vorria et al., 2003; Zeanah et al., 2005).

An early descriptive study that followed children after they had been removed from an institution and placed in foster care was conducted by Goldfarb (1943). He described 10- to 14-year-old children who had lived in institutions for their first 3 years of life as withdrawn and removed from their foster parents and caseworkers. He concluded that children who had been institutionalized in the early years of life could not form attachments.

A subsequent study with improved methodology reached a different conclusion. Tizard and her colleagues (Hodges & Tizard, 1989; Tizard & Hodges, 1978; Tizard & Rees, 1974, 1975) studied 65 children abandoned at birth due to poverty and placed in residential nurseries in London. Between ages 2 and 4 years, 15 of the children were returned to their biological parents, 24 were adopted, and 26 were still being reared in the residential nurseries. At age 4 years, the children were evaluated. Although the majority of the 26 children who remained in institutions at age 4 years showed serious disturbances of attachment, this was not true for either the adopted or the returned groups. In fact, 80% of the adoptive parents reported that their children had become attached to them within 1 year, and the majority reported that the children were deeply attached. Nevertheless, even in the adoptive group, which had a more favorable outcome than the other two groups across all measures and ages studied, some children displayed signs of indiscriminate behavior, that is, lack of reticence about approaching, interacting with, and going off with a stranger. None of the comparison group of children who had never been institutionalized showed indiscriminate behavior. Thus, Tizard's findings suggested more plasticity in attachment than Goldfarb (1943) had believed possible.

MacLean (2003) has pointed out correctly that one reason for the discrepancy may be that the children Goldfarb (1943) studied experienced more severe privation than the children Tizard and her colleagues (Hodges & Tizard, 1989; Tizard & Hodges, 1978; Tizard & Rees, 1974, 1975) studied. This was reflected both in descriptions of the institutional caregiving practices and in the IQs of the two groups. The children in the Tizard study who lived in residential nurseries had IQs in the average range, whereas the children in the Goldfarb study were delayed in all domains. Nevertheless, the main problem with these studies is that the measures of attachment they used were neither standardized nor derived from behavioral observations. More recent studies have addressed these shortcomings by studying attachment in preschool children.

Marcovitch et al. (1997) compared fifty-six 3- to 5-year-old Canadian children who had been adopted out of Romanian institutions before 6 months of age to those who had been adopted after 6 months of age, as well as a group of never-institutionalized Canadian comparison children. They found that secure attachment was observed less often (30%) controlling/insecure-other more often (42%).

Chisholm (1998) studied children adopted into Canada after having spent at least 8 months in a Romanian institution (*n* = 46), children adopted into Canada from Romania at < 4 months of age (*n* = 30), and a Canadian-born comparison group matched on age and sex to the first group (*n* = 46). Attachment at 39 months postadoption was assessed with a home-based version of the SSP and coded with the Preschool Assessment of Attachment (Crittenden, 1992). Children who had spent 8 months or more in institutions were less likely to be securely attached than the other two groups of children, 37% versus 66% of those adopted after 4 months or less versus 58% of the Canadian children. In addition, there were more atypical insecure attachments in the group who had spent 8 months or more in institutions, 32.6% versus 3.7% of the < 4 months group and 6.9% of the Canadian children.

O'Connor, Marvin, Rutter, Olrick, and Britner (2003) studied attachment, using a within-the-home SSP in 111 four-year-old children adopted out of Romanian institutions into the United Kingdom and compared them to 52 nonmaltreated children adopted within the United Kingdom at < 6 months of age. They found that 55.1% of the children adopted within the United Kingdom were securely attached, with 4.1% disorganized-controlling and 16.3% insecure-other. In contrast, the Romanian children adopted out of institutions prior to 6 months of age were 41% secure, 9.8% disorganized-controlling, and 38.5% insecure-other. Of the children adopted after 6 months from institutions, only 33.3% were secure, whereas 7.7% were disorganized-controlling and 51.7% were insecure-other.

Taken together, these results indicate that children do form attachments following early deprivation, but subsequent attachment relationships are less likely to be secure and more likely to be atypical (disorganized-controlling or insecure-other). Organized insecure patterns of attachment (avoidant, ambivalent, and dependent) do not appear to be increased in children with institutional rearing.

Adoption studies such as these, informative as they are, also have limitations. These studies include only children who were selected intentionally rather than randomly, and they involve changes for the child not only from institutional care to family care but also changes in spoken language and culture.

In part to address these limitations, we initiated the Bucharest Early Intervention Project (BEIP; Zeanah et al., 2003). This is the first-ever randomized controlled trial of foster care as an alternative to institutional care. The BEIP was begun in the fall of 2000, involving 136 children living in institutions in Bucharest, Romania. They were recruited between the ages of 6 and 31 months, assessed comprehensively, and then assigned at random to foster care or to continued institutional care. They were reassessed at 30, 42, and 54 months of age, and the children in foster care were compared to children who received continued institutional care and to a group of never-institutionalized children who were recruited from pediatric clinics and living with their families. In this report, we report on differences in attachment in the children at 42 months.

Specifically, we examined the following questions. Did placement of young children being raised in institutions into families increase security of attachment and decrease atypical patterns of attachment at 42 months of age compared to children who received ongoing institutional care? Did age at placement in foster care affect the likelihood of children developing secure or organized attachments at 42 months of age? Did factors such gender, ethnicity, developmental quotient (DQ), or age at placement moderate the effects of intervention on security or organization of attachment?

Method

Participants

Three groups of children participated in this study. They were all participants in the BEIP, a longitudinal study aimed at examining recovery from early deprivation (Zeanah et al., 2003) through placement in foster care as an intervention to address the effects of early institutionalization. All of the children < 31 months of age who were cared for in the six institutional settings in Bucharest, Romania at the beginning of the study ($n = 187$) were assessed for participation in the BEIP. Exclusion criteria for the study included medical conditions such as genetic syndromes, signs of fetal alcohol syndrome, and microcephaly. A total of 51 children were excluded from the study. The sample consisted of children ($n = 136$) who had been abandoned very early in their lives and placed in institutions. After extensive baseline assessments (see Smyke et al., 2007, for a full description), the 136 institutionalized children were randomized into the continued care as usual including institutionalization group (care as usual group [CAUG]; ($n = 68$) or to the BEIP foster care group (FCG; $n = 68$). Randomization was accomplished by assigning a number to each of the 136 study children. These numbers (1–136) were placed in a hat and then drawn at random from the hat. The first number drawn was assigned to the CAUG and the second number was assigned to the FCG, with this pattern repeated until all 136 numbers had been drawn. There were two sets of twins in the study. The twins' numbers were placed on the same piece of paper and each set of twins was placed together.

A third group of children comprised children reared in their families and recruited from pediatric clinics in Bucharest (NIG; $n = 72$). This group was not randomly assigned and served as a comparison group. Figure 1 illustrates the flow of participants through the study. All children were originally recruited when they were between 6 and 31 months of age. Data presented in this study were obtained during the course of the 42-month assessment, which occurred 11–36 months after randomization. As noted in Figure 1, there were 57 CAUG, 61 FCG, and 51 NIG who completed SSPs with usable data at 42 months.

For ethical and legal reasons, we did not interfere with any decisions regarding children's placements, leaving those decisions to Romanian child protection officials. The one exception was that we ensured that no child randomized to foster care would be readmitted to institutional care during or after the study. Between the baseline and 42-month assessments, a number of children changed status (see Figure 1). In the following analyses, we followed an "intent-to-treat" approach, so that children were assessed in their original group (CAUG or FCG), regardless of their status at 42 months.

Additional details regarding sample recruitment are available (Zeanah et al., 2005). Table 1 provides demographic information for the study sample. As noted in Table 1, the ethnic distribution of the randomized groups (CAUG and FCG) is similar (Roma ~33%) but it differs from the comparison group. The overrepresentation of children of Roma heritage in institutions has been noted previously (Stephenson et al., 1993) and probably reflects the impoverished circumstances of many Roma. Early reports suggested that almost half of the young children in Romanian institutions were of Roma heritage (Stephenson et al., 1993), despite the estimates that Roma account for < 10% of the population of Romania.

BEIP Foster Care Intervention

Foster care was conceptualized as an intervention to address the effects of institutionalization on young children. When the project began in 2000, there were almost no foster homes in Bucharest, and they primarily had been used by adoption agencies to prepare children for the transition from institution to adoption. BEIP created a network of 56 foster homes in Bucharest. Foster parents were recruited, completed a background check, and trained by Romanian trainers using a manual derived from similar manuals used in the United States. Foster parents were considered full-time employees and received a salary and benefits. This was in contrast to the United States, where foster parents usually receive board payment for each child and do not receive payment if no child is in their home.

Foster parents received support from and were monitored by BEIP social workers who visited with them regularly. Social workers assisted foster parents in managing challenging postinstitutional behavior and organized a support group for foster parents early in the project. Importantly, social workers supported the foster parents in the establishment of warm, supportive, and committed relationships with their young foster children. BEIP foster care was designed to be "child centered,"

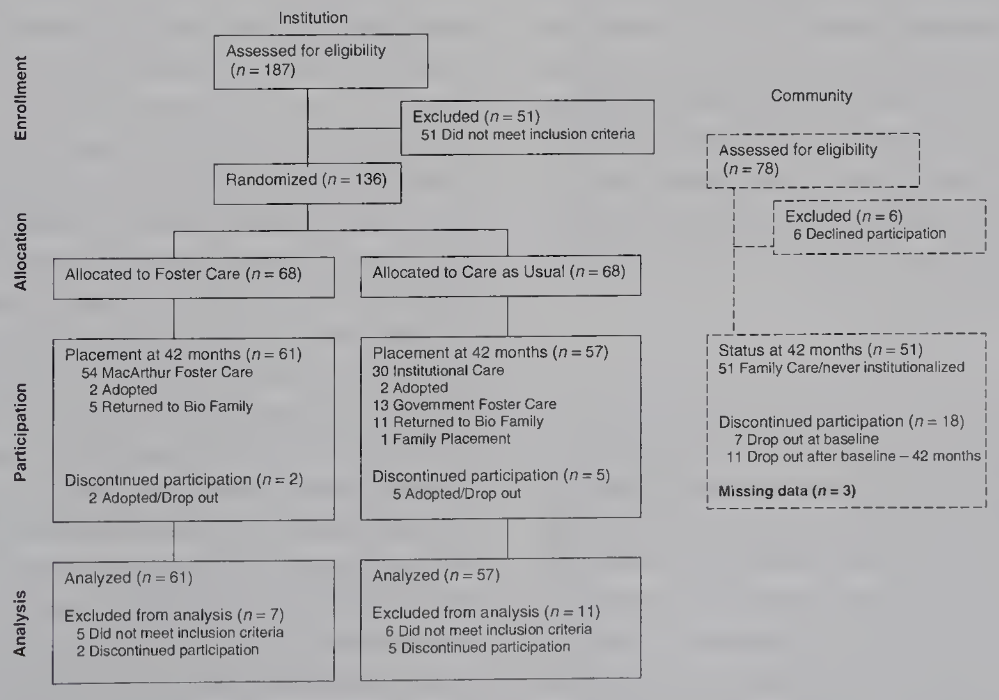

Institution
Enrollment
Assessed for eligibility ($n = 187$)
Excluded ($n = 51$)
51 Did not meet inclusion criteria
Randomized ($n = 136$)
Allocation
Allocated to Foster Care ($n = 68$)
Allocated to Care as Usual ($n = 68$)
Participation
Placement at 42 months ($n = 61$)
54 MacArthur Foster Care
2 Adopted
5 Returned to Bio Family
Discontinued participation ($n = 2$)
2 Adopted/Drop out
Placement at 42 months ($n = 57$)
30 Institutional Care
2 Adopted
13 Government Foster Care
11 Returned to Bio Family
1 Family Placement
Discontinued participation ($n = 5$)
5 Adopted/Drop out
Analysis
Analyzed ($n = 61$)
Excluded from analysis ($n = 7$)
5 Did not meet inclusion criteria
2 Discontinued participation
Analyzed ($n = 57$)
Excluded from analysis ($n = 11$)
6 Did not meet inclusion criteria
5 Discontinued participation
Community
Assessed for eligibility ($n = 78$)
Excluded ($n = 6$)
6 Declined participation
Status at 42 months ($n = 51$)
51 Family Care/never institutionalized
Discontinued participation ($n = 18$)
7 Drop out at baseline
11 Drop out after baseline – 42 months
Missing data ($n = 3$)

Figure 1. Flow of participants through the Bucharest Early Intervention Project study: Institution (randomized) and community (not randomized).

Table 1
Demographic Characteristics of the Study Sample

Child characteristics	Care as usual ($n = 57$)	Foster care ($n = 61$)	Never institutionalized ($n = 51$)	Significance
Age in months (*SD*)	42.44 (.39)	42.37 (.28)	43.15 (.98)	NIG > CAUG, FCG***
Ethnicity, *n* (%)				
Romanian	26 (45.6)	35 (57.4)	46 (90.2)	NIG ≠ CAUG, FCG***
Roma (Gypsy)	21 (36.8)	18 (29.5)	4 (7.8)	
Unknown/other	10 (17.5)	8 (13.1)	1 (2.0)	
Gender				
Male	49.1%	52.5%	47.1%	NIG = CAUG = FCG

Note. CAUG = care as usual group; FCG = foster care group; NIG = never-institutionalized group.
***$p < .001$.

that is, focused on recognizing the psychological as well as the instrumental needs of the young children in their care. In contrast to the frequent placement changes sometimes seen in foster care in the United States, there were very few moves within BEIP foster care. Several of the foster parents ultimately adopted the children they cared for, despite the fact that they would then no longer receive a salary for caring for that child. Foster parents were encouraged by project social workers to accept their foster children as members of the family and to become the "psychological parents" of the children for whom they cared.

The social workers were supported throughout the project by clinicians in the United States who provided weekly phone or video consultations. The consultants also visited the study center quarterly to provide in-person consultation, visit foster homes, and discuss issues that had arisen in the care of this challenging group of children. The emphasis was on developing attachment relationships between foster parents and children, facilitating language development in children, and providing foster parents with techniques for managing difficult child behavior (see Nelson et al., 2007, Supporting Online Material, for further information about the intervention).

Measures

Attachment. The quality of children's attachment to their caregivers was assessed using Ainsworth's SSP (Ainsworth et al., 1978). Community children were seen with their mothers, foster children with their foster mothers, and institutionalized children were seen with their "favorite" caregivers, as identified by staff. If the child did not have an identified favorite, they were seen with a caregiver who worked with them regularly and knew them well.

Coders, who were native Romanians, were trained to reliability in the MacArthur Preschool System with W. Whelan (Cassidy & Marvin, with the MacArthur Working Group, 1992). All procedures were coded by a primary coder who was blind to the children's group status. In addition, 120 (75%) of the procedures were double-coded to reliability; on the five-way (secure, avoidant, ambivalent, disorganized-controlling, and insecure-other) classification, kappa = 0.81 (95% CI = 0.73, 0.89). Percent agreement was 87.7%. For secure versus not secure coding, kappa = 0.83 (95% CI = 0.73, 0.93), percent agreement = 91.7%. For organized (secure, avoidant, and ambivalent) versus atypical (disorganized, controlling, and insecure-other) classifications, kappa = 0.93 (95% CI = 0.85, 1.01), percent agreement = 97.5%. Disagreements were resolved by conferencing.

In keeping with the MacArthur coding system, categories coded included secure (B), avoidant (A), ambivalent-dependent (C), disorganized-controlling (D), and insecure-other (I-O). Although the underlying strategy for obtaining comfort and reassurance from the attachment figure in times of distress remains the same as that seen in infancy, different behaviors are evident in preschool children. For example, secure attachment involves demonstrating a positive, engaged and open style of verbal and nonverbal interaction but less proximity seeking than in infancy. Children whose attachment is classified as avoidant have reduced involvement with the caregiver and do not seek contact when distressed. In addition, they say little to the caregiver, particularly about feelings, and they remain affectively neutral. Children whose attachment is ambivalent-dependent display passivity, helplessness, immaturity, or petulance and resistance in their interactions with the caregiver. Some preschool children whose attachment is classified as disorganized display behaviors similar to toddlers classified as disorganized (see Main & Solomon, 1990), but others exhibit efforts to control the behavior of the caregiver, particularly following the stressful separation episodes of the SSP. Such behavior tends to follow one of two patterns: controlling-caregiving or controlling-punitive (Cassidy & Marvin, with the MacArthur Working Group, 1992). Because disorganized and controlling behavior are conceptually linked, the classification is called disorganized-controlling. Finally, Marvin and his colleagues (Cassidy & Marvin, with the MacArthur Working Group, 1992) identified an ''insecure-other'' classification to describe children who did not fit into the traditionally insecure categories. This classification was prevalent in previous studies of formerly institutionalized children (Marcovitch et al., 1997; O'Connor et al., 2003). Children who are classified as insecure-other are clearly not secure, but they also do not fit into one of the insecure categories described earlier. For example, some children classified as insecure-other display a fearful, ''compulsive-compliant'' pattern; others are severely emotionally dysregulated in the context of their caregiver. Still others exhibit both avoidant and resistant behaviors in their interactions with their caregivers. In other words, such children have not evolved a reliable strategy for managing their distress by seeking and receiving comfort and reassurance from their caregivers. Although behavior in interaction with their caregivers may vary, the underlying attachment representation for children who are categorized controlling or insecure-other, may be characterized by ''themes of fear, confusion, chaos, and disorganization'' (Teti, 1999, p. 230).

In addition to assigning a best fitting classification of attachment, coders also assigned a continuous rating of security, using a scale from 1 to 9, with 1 = *no security evident* to 9 = *most secure* (Cassidy & Marvin, with the MacArthur Working Group, 1992). Reliability for this coding was excellent ($r = .87$).

Cognitive development. The Bayley Scales of Infant Development II (BSID–II; Bayley, 1993) were used to assess developmental status. The Mental Development Index (MDI), a scaled score, ranged from 50 to 150. Children who obtained scaled scores below 50 were assigned a numeric MDI score of 49. Raw scores were assigned an age equivalent score to enable analyses when scaled scores < 50 were obtained (Lindsey & Brouwers, 1999). DQs were computed for each child (i.e., [age equivalent score/chronological age] × 100).

Quality of caregiving. The Observational Record of the Caregiving Environment (National Institute of Child Health and Human Development Early Child Care Research Network, 1996, 1997) was adapted for use in assessing the quality of caregiving for a given child in his or her environment (see Zeanah et al., 2005, for an in-depth description of this measure). Children were videotaped in their ''home'' environment (i.e., CAUG: institution; FCG: foster home; NIG: family home) for 1½ hr with their favorite caregiver. Coders watched the videotape and coded qualitative items (e.g., sensitivity, intrusiveness) on a scale that ranged from 1 = *not at all characteristic* to 4 = *highly characteristic.*

A quality-of-caregiving score was determined by calculating the mean of the scores for sensitivity, stimulation of development, positive regard for child, flat affect (reversed), and detachment (reversed). Possible quality-of-caregiving scores ranged from 1 to 4. Scale reliability was acceptable (Cronbach's $\alpha = .82$).

Consent

Following approvals by the Institutional Review Boards of the three principal investigators (Zeanah, Nelson, and Fox) and by the local Commissions on Child Protection in Bucharest, the study began in collaboration with the Institute of Maternal and Child Health of the Romanian Ministry of Health. Consent was signed by the Commissioner for Child Protection for each child participant living in his or her sector of Bucharest, as dictated by Romanian law. Further assent for each procedure was obtained from each caregiver who accompanied the child to the visit.

Ethical issues involving this vulnerable population have been discussed in detail elsewhere (Millum & Emanuel, 2007; Nelson et al., 2007; Wassenaar, 2006; Zeanah et al., 2006).

Results

Overall distribution of attachment classifications, organization, and security, first, are presented by group. Then we compare security ratings for the groups. We then examine the important question of whether there is a critical age after which placement in foster care is less likely to result in an organized or secure attachment relationship. Finally, we examine whether specific factors such as gender, quality of caregiving, or cognitive skills are useful for predicting organization

Table 2
Distribution of Strange Situation Procedure Classifications

Strange Situation procedure classification	Care as usual (%) (n = 57)	Foster care (%) (n = 61)	Never institutionalized (%) (n = 51)
Secure (B)	10 (17.5)	30 (49.2)	33 (64.7)
Avoidant (A)	14 (24.6)	12 (19.7)	6 (11.8)
Ambivalent (C)	7 (12.3)	5 (8.2)	7 (13.7)
Disorganized-controlling (D)	3 (5.3)	8 (13.1)	5 (9.8)
Insecure-other (I-O)	23 (40.4)	6 (9.8)	0
Organized (A, B, C)	31 (54.4)	47 (77.0)	46 (90.2)
Atypical (D-controlling, I-O)	26 (45.6)	14 (23.0)	5 (9.8)

or security in young children previously institutionalized.

SSP Classifications

The frequency distributions of SSP classifications are shown in Table 2. Differences were most apparent at the extremes of the classifications. Thus, only 17.5% of the CAUG but 49.5% of the FCG were securely attached. Further, 40% of the CAUG but only 13.1% of the FCG were classified as insecure-other. The distribution of the classifications across the three groups differed significantly, $\chi^2(8) = 48.30$, $p < .001$. Comparison of the CAUG and FCG showed a significant difference, $\chi^2(4) = 22.62$, $p < .001$, but no difference was found between FCG and the NIG, $\chi^2(4) = 8.34$, $p = .08$. No significant differences were found between males and females in overall distribution of attachment classifications, nor were gender differences noted in the distributions of any of the individual groups. Ethnicity also was unrelated to any of the attachment measures.

Organization of attachment was dichotomized organized (A, B, and C) and atypical (D-controlling and I-O) as shown in Table 2. Significant differences were found among the three groups, $\chi^2(2) = 18.33$, $p < .001$, and between the CAUG and FCG, $\chi^2(1) = 6.75$, $p < .01$, but not between the FCG and the NIG, $\chi^2(1) = 3.41$, $p = .065$. Differences were noted between males and females in organized versus atypical status for the FCG only, $\chi^2(1) = 4.97$, $p < .05$, with FCG girls more likely to be organized than FCG boys.

In an even more stringent test of the intervention, we compared the frequency of secure (B) versus not secure (all others). Significant differences were found among the three, $\chi^2(2) = 25.79$, $p < .001$, and between CAUG and FCG, $\chi^2(1) = 13.16$, $p < .001$, but not between the FCG and the NIG, $\chi^2(1) = 2.72$, $p = .099$. Gender differences were found within the FCG, $\chi^2(1) = 3.67$, $p = .055$, and the NIG, $\chi^2(1) = 4.29$, $p = .038$; girls were more often secure than boys within this group.

Security Ratings

Coders assigned security ratings ranging from 1 to 9 for each procedure. Table 3 shows the average security rating for each group overall and by gender. Notably, even among dyads classified as secure in the CAUG, security ratings were lower than those of either the FCG or the NIG. Main effects of group, $F(2, 169) = 17.10$, $p < .001$, and of secure versus insecure, $F(1, 169) = 386.37$, $p < .001$, were significant. This pattern was additive across groups, $F(2, 169) = 0.043$, $p = .958$. Security scores were highest among the NIG, next highest among the FCG, and lowest in the CAUG. A similar pattern was present for the organized versus atypical analyses. There was no association of these scores with gender or ethnicity.

Effects of Age at Placement on Organization and Security

To investigate the effect of age at placement into foster care on organization and security, FCG children were grouped by dichotomous age at

Table 3
Security Rating by Group

	Care as usual (%) (n = 57)	Foster care (%) (n = 61)	Never institutionalized (%) (n = 51)
Security rating	3.13 (1.33)	4.61 (1.64)	5.49 (1.52)
Male	3.13 (1.39)	4.29 (1.59)	5.09 (1.65)
Female	3.13 (1.40)	4.96 (1.64)	5.83 (1.34)

Note. Security ratings: 1 = *no security evident*, 9 = *most secure.*

Table 4
Organized Versus Atypical by Placement Age Group

Age cutoff (months)	Percent organized (SSP)			
	Younger	Older	$\chi^2(1)$	*p* Value
18	100	70.2	5.05	.025
20	100	67.4	7.22	.007
22	100	63.2	10.57	.001
24	96.6	58.1	12.41	.000
26	90.9	59.3	8.32	.004
28	85.0	60.0	4.66	.031

Note. SSP = Strange Situation procedure.

Table 5
Secure Versus Insecure by Placement Age Group

Age cutoff (months)	Percent secure (SSP)			
	Younger	Older	$\chi^2(1)$	*p* value
18	53.8	46.8		*ns*
20	58.8	44.2		*ns*
22	63.6	39.5	3.26	.071
24	69.0	29.0	9.57	.002
26	66.7	25.9	9.87	.002
28	60.0	25.0	6.54	.011

Note. SSP = Strange Situation procedure.

placement in foster care (younger than cutoff vs. older than cutoff) for placement cutoffs of 18, 20, 22, 24, 26, and 28 months and then compared by organized versus atypical status and by secure versus insecure status. Tables 4 and 5 show significant differences between organized versus atypical status for children placed earlier when compared to those placed later for all age cutoffs. For secure versus insecure status, however, a deflection point was evident between 22 and 24 months, suggesting that children placed before 24 months had a better response to the intervention of foster care.

Predicting Attachment Organization and Security

Multiple logistic regression was used to determine whether a child had an organized attachment (as opposed to atypical [disorganized or insecure-other] classification) at 42 months of age. Predictors included gender, quality of caregiving, *M* (*SD*): CAUG: 2.54 (0.58); FCG: 2.78 (0.57); NIG: 2.73 (0.57); FCG > IG, and cognitive development (Nelson et al., 2007; see Table 6) as predictors. Logistic regression analyses were conducted for the CAUG, FCG, and NIG separately. For the CAUG, neither caregiving quality nor gender significantly predicted attachment organization, but cognitive development did (Wald $\chi^2 = 5.54$, $p = .019$, $df = 1$). For the children in the CAUG, an increase of 1 unit in DQ at 42 months of age was associated with an increase in the odds of a child receiving care as usual having an organized attachment by a factor of 1.072 (95% CI = 1.012, 1.135). Similar logistic regression analyses performed on the never-institutionalized group and the FCG showed no significant associations between the odds of a child developing an organized attachment with his or her caregiver based on gender, cognitive development, or quality of caregiving. However, when length of time in foster care was added to the logistic regression for the FCG, after controlling for the effect of the other factors, it was the sole significant factor (Wald $\chi^2 = 4.94$, $p = .026$, $df = 1$). An increase of 1 month in length of time in foster care was associated with an increase in the odds of a child in the FCG having an organized attachment by a factor of 1.273 (95% CI = 1.029, 1.574). Length of time in foster care was examined for the FCG only.

Similar analyses were conducted, using gender, quality of caregiving, and cognitive development as predictors to determine whether a child had a secure attachment (as opposed to an insecure attachment classification) at 42 months of age. Logistic regression analyses again were conducted for the CAUG, FCG, and NIG separately. For the

Table 6
Developmental Quotient by Organized Versus Atypical and by Secure Versus Insecure Status

	Care as usual group (%) (*n* = 56)	Foster care group (%) (*n* = 61)	Never-institutionalized group (%) (*n* = 51)	
Developmental quotient	76.90 (13.31)	85.49 (14.23)	103.30 (11.85)	NIG > FCG > CAUG
Organized	81.36 (10.30)	87.97 (13.11)	104.50 (11.49)	Organized > atypical
Atypical	71.36 (14.68)	77.32 (15.18)	92.51 (10.38)	
Secure	77.40 (10.66)	91.03 (11.29)	107.48 (7.50)	Secure > insecure
Insecure	76.79 (13.92)	80.30 (14.89)	95.88 (14.54)	

CAUG, neither caregiving quality, gender, nor cognitive development was associated with secure versus insecure status. However, for the FCG and the NIG, cognitive development at 42 months of age was associated with an increase in the odds of a child having a secure attachment (FCG: Wald $\chi^2 = 5.022$, $p = .022$, $df = 1$; NIG: Wald $\chi^2 = 7.247$, $p = .007$, $df = 1$). An increase of 1 unit in DQ was associated with an increase in the odds of a child in the FCG having a secure attachment by a factor of 1.056 (95% CI = 1.008, 1.106). For children in the NIG, a 1-unit increase in DQ was associated with an increase in the odds of a child in that group having a secure attachment by a factor of 1.144 (95% CI = 1.037, 1.262). When length of time in foster care was added to the logistic regression for the FCG, cognitive development became marginally significant (Wald $\chi^2 = 3.71$, $p = .054$, $df = 1$) but the other factors were not associated with increased odds of secure attachment.

Discussion

This study reports attachment results from the first ever randomized trial of foster care as an intervention for young children who had been abandoned and placed in institutions. We had earlier reported that in this sample, at a mean age of 21 months, institutional care was associated with profound disturbances of attachment, as reflected in part by 65% of the children classified as disorganized and another 13% who showed so little attachment behavior that they were designated "unclassified" in the SSP (Zeanah et al., 2005). Following placement in care for 12–36 months ($M = 19.12$ months, $SD = 6.80$ months), children in foster care made significant gains compared to children who remained in institutions. This was reflected in significantly more of them having secure attachment classifications and also having higher ratings of security. Further, significantly fewer of them had atypical (i.e., disorganized-controlling or insecure-other) classifications. Because we used intent-to-treat analyses, the findings we report are likely conservative estimates of intervention effects.

These findings indicate that young children placed into foster care after early institutional rearing may experience significant recovery with regard to attachment. Furthermore, because of the randomized design, we can conclude that the placement in families was causally related to improvement in the children's attachment status. This is one of several studies that have shown that changing caregiving environments changes children's attachments (Cicchetti et al., 2006; Dozier et al., 2007; Fisher & Kim, 2007; Heinicke et al., 1999; Klein Velderman et al., 2006; Toth et al., 2002; van den Boom, 1994).

The magnitude of these results is notable. For example, the 49% of children classified securely attached in the FCG is greater or equal to what has been reported in two studies of children adopted internationally out of Romanian institutions. Marcovitch et al. (1997) found 30% of 3- to 5-year-olds and O'Connor et al. (2003) found 47% of 4-year-olds were securely attached to their adoptive mothers. Even more striking, only 13% of the FCG were classified as insecure-other, whereas 42% of the Marcovitch et al. (1997) sample and 47% of the O'Connor et al. (2003) sample had this unusual classification. Although adoptive parents in Canada and the United Kingdom are likely to have socioeconomic advantages not available to foster parents in Romania, this may be overshadowed by the absence of cultural transition for the BEIP children compared to those adopted internationally. Because both the Marcovitch and O'Connor samples included children adopted soon after the fall of Ceausescu, the differences in attachment following placement in families may result from more severe conditions of privation than existed 10 years later when the BEIP was conducted. For example, the largest institution for young children in Romania, St. Catherine's Placement Center, known as Bucharest Number One during the Communist era, had 850 children living there in 1989, but only 450 in the fall of 2000.

A second important finding was that age at which children were placed in foster care was powerfully related to recovery of attachment. This was evident in both analyses of organized attachments (secure, avoidant, and ambivalent) and of secure attachment. The younger a child was placed in foster care, the more likely the child would develop an organized attachment at 42 months. For security of attachment, however, children placed before 24 months of age were more likely to have secure attachments at 42 months. One implication of these findings is that at least in the first 3 years of life children retain considerable plasticity for adaptation with regard to attachment. This supports our theoretical notion that attachment reflects experience-expectant neural processes. That is, even following experiences of social and material deprivation, children who are placed in more species typical caregiving environments exhibit a strong disposition to form attachments. On the other hand, another implication of the findings is a decreasing plasticity of attachment with increasing age and are in keeping with Dozier and Bick's (2007) observation that children placed in foster care younger than 12 months were able to develop trusting relationships with new caregivers more quickly than older infants and toddlers (Dozier & Bick, 2007). An important question for future research is the degree to which subsequent experiences can compensate for early adversity and specifically whether the large number of insecure-other attachment classifications predicts subsequent psychopathology or interpersonal difficulties.

Gender seemed to moderate the effects of the intervention, with girls showing more response to the foster care intervention than boys. The pattern of boys in high-risk samples showing greater incidence of disorganized attachment than girls has been noted by Carlson, Cicchetti, Barnett, and Braunwald (1989) and others (David & Lyons-Ruth, 2005; Vondra, Hommerding, & Shaw, 1999). In contrast, most studies examining attachment organization, particularly in lower risk samples, do not show gender differences. It may be that male gender, in the context of markedly compromised environments, poses a risk that it does not in less challenging environments.

On the other hand, we noted that when gender was included in the regression analysis, along with caregiving quality and cognitive development, as factors influencing the likelihood of developing an organized attachment relationship or a secure attachment relationship, it was not significant for any of the groups. In fact, caregiving quality was also not significant. Rather, for the CAUG, greater cognitive development suggested greater likelihood of developing an organized relationship but not of developing a secure one. It is possible that in the context of the institutional setting, children more capable of engaging their caregivers are more likely to develop an organized attachment with their caregivers, although not a secure one. Given that age of placement is confounded with length of time in foster care, it is not unexpected that this factor is an important influence in the development of organized attachment for the FCG.

Individual differences in caregiving quality at 42 months were not related to children's attachment organization. This may be a result of the single, relatively brief observation being insufficient as a measure of caregiving quality, especially considering that a number of children had changed caregiving environments between baseline and 42 months.

The present study is not without limitations. First, because of the practice in Romania of keeping abandoned children in maternity hospitals during the first 6 months of life, none of the children we studied were younger than 6 months of age. This makes evaluating the impact of very early foster care impossible. In addition, comparisons with studies of children adopted out of institutions are more challenging because both the Rutter et al. (2007) and Chisholm (1998) studies found more favorable development in children adopted younger than 6 months. We chose to use the MacArthur approach to code our SSPs and we recognize that whereas several studies have used this methodology, others have used the Preschool Assessment of Attachment (Crittenden, 1992). A comparison of these two methods might be of interest but was beyond the scope of this study. Second, as seen in Figure 1, despite initial random assignment of the institutionalized children to either the CAUG or the FCG, a great deal of movement between and among groups occurred. This was an important ethical constraint, but it made pure group comparisons impossible. We used intent-to-treat analysis and study of the figure indicates that the majority of the movement was in children assigned to CAUG who ended up in families, either adoptive or their own biological families. As a result, we believe that the results we reported are a conservative estimate of the impacts of the intervention. Finally, as noted earlier, one reason we were unable to demonstrate dose–response effects of the intervention may be because the measure of caregiving quality that we used was based on too brief a period of observation or insensitive to important differences.

The unique population of children reared in institutions remains important in our efforts to understand more fully the effects of atypical environments on young children's development and the processes through which they recover from early adversity.

References

Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the Strange Situation*. Hillsdale, NJ: Erlbaum.

Bayley, N. (1993). *Bayley Scales of Infant Development* (2nd ed.). New York: Psychological Corporation.

Boris, N. W., Hinshaw-Fuselier, S. S., Smyke, A. T., Scheeringa, M., Heller, S. S., & Zeanah, C. H. (2004).

Comparing criteria for attachment disorders: Establishing reliability and validity in high-risk sample. *Journal of the American Academy of Child and Adolescent Psychiatry, 43*, 568–577.

Carlson, V., Cicchetti, D., Barnett, D., & Braunwald, K. (1989). Disorganized/disoriented attachment relationships in maltreated infants. *Developmental Psychology, 25*, 525–531.

Cassidy, J., & Marvin, R. S., with the MacArthur Working Group (1992). *Attachment organization in preschool children: Procedures and coding manual* (4th ed.). Unpublished manuscript, University of Virginia.

Chisholm, K. (1998). A three year follow-up of attachment and indiscriminate friendliness in children adopted from Romanian orphanages. *Child Development, 69*, 1092–1106.

Cicchetti, D., Rogosch, F. A., & Toth, S. L. (2006). Fostering secure attachment in infants in maltreating families through preventive interventions. *Development and Psychopathology, 18*, 623–649.

Crittenden, P. M. (1992). Quality of attachment in the preschool years. *Development and Psychopathology, 4*, 209–241.

David, D. H., & Lyons-Ruth, K. (2005). Differential attachment responses of male and female infants to frightening maternal behavior: Tend or befriend versus fight or flight? *Infant Mental Health Journal, 26*, 1–18.

De Wolff, M. S., & van IJzendoorn, M. H. (1997). Sensitivity and attachment: A meta-analysis on parental antecedents of infant attachment. *Child Development, 68*, 571–591.

Dozier, M., & Bick, J. (2007). Changing caregivers: Coping with early adversity. *Psychiatric Annals, 37*, 411–415.

Dozier, M., Grasso, D., & Lindhiem, O. (2007). The role of caregiver commitment in foster care: Insights from the This Is My Baby Interview. In D. Oppenheim & D. Goldsmith (Eds.), *Attachment theory in clinical work with children: Bridging the gap between research and practice* (pp. 90–108). New York: Guilford.

Fisher, P. A., & Kim, H. Y. (2007). Intervention effects on foster preschoolers' attachment-related behaviors from a randomized trial. *Prevention Science, 8*, 161–170.

Goldfarb, W. (1943). The effects of early institutional care on adolescent personality. *Journal of Experimental Education, 12*, 106–129.

Greenough, W. T., Black, J. E., & Wallace, C. S. (1987). Experience and brain development. *Child Development, 58*, 539–559.

Heinicke, C. M., Fineman, N. R., & Ruth, G. (1999). Relationship-based intervention with at-risk mothers: Outcome in the first year of life. *Infant Mental Health Journal, 20*, 349–374.

Hodges, J., & Tizard, B. (1989). Social and family relationships of ex-institutional adolescents. *Journal of Child Psychology and Psychiatry, 30*, 77–97.

Klein Velderman, M., Bakermans-Kranenburg, M. J., & Juffer, F. (2006). Effects of attachment-based interventions on maternal sensitivity and infant attachment: Differential susceptibility of highly reactive infants. *Infant Mental Health Journal, 27*, 466–493.

Koren-Karie, N., Oppenheim, D., & Doley, S. (2002). Mother's insightfulness regarding their infants' internal experience: Relations with maternal sensitivity and infant attachment. *Developmental Psychology, 38*, 534–542.

Lindsey, J. C., & Brouwers, P. (1999). Intrapolation and extrapolation of age-equivalent scores for the Bayley II: A comparison of two methods of estimation. *Clinical Neuropharmacology, 22*, 44–53.

MacLean, K. (2003). The impact of institutionalization on child development. *Development and Psychopathology, 15*, 853–884.

Main, M., & Solomon, J. (1990). Procedures for identifying infants as disorganized/disoriented during the Ainsworth Strange Situation. In M. T. Greenberg, D. Cicchetti, & E. M. Cummings (Eds.), *Attachment in the preschool years: Theory, research, and intervention* (pp. 121–160). Chicago: University of Chicago Press.

Marcovitch, S., Goldberg, S., Gold, A., Washington, J., Wasson, C., Krekewich, K., et al. (1997). Determinants of behavioural problems in Romanian children adopted into Ontario. *International Journal of Behavioral Development, 20*, 17–31.

Millum, J., & Emanuel, E. J. (2007). The ethics of international research with abandoned children. *Science, 318*, 1874–1875.

National Institute of Child Health and Human Development Early Child Care Research Network. (1996). Characteristics of infant child care: Factors contributing to positive caregiving. *Early Childhood Research Quarterly, 11*, 269–306.

National Institute of Child Health and Human Development Early Child Care Research Network. (1997). The effects of infant child care on infant–mother attachment security: Results of the NICHD Study of Early Child Care. *Child Development, 68*, 860–879.

Nelson, C. A., Zeanah, C. H., Fox, N. A., Marshall, P. J., Smyke, A. T., & Guthrie, D. (2007). Cognitive recovery in socially deprived young children: The Bucharest Early Intervention Project. *Science, 318*, 1937–1940.

O'Connor, T. G., Marvin, R. S., Rutter, M., Olrick, J. T., & Britner, P. A. (2003). Child–parent attachment following early institutional deprivation. *Development and Psychopathology, 15*, 19–38.

Rutter, M., Colvert, E., Kreppner, J., Beckett, C., Castle, J., Groothues, C., et al. (2007). Early adolescent outcomes for institutionally-deprived and non-deprived adoptees. I: Disinhibited attachment. *Journal of Child Psychology and Psychiatry, 48*, 17–30.

Slade, A., Sadler, L. S., & Mayes, L. C. (2005). Minding the baby: Enhancing parental reflective functioning in a nursing/mental health home visiting program. In L. J. Berlin, Y. Ziv, L. Amaya-Jackson, & M. T. Greenberg (Eds.), *Enhancing early attachments: Theory, research, intervention, and policy* (pp. 152–177). New York: Guilford.

Smyke, A. T., Dumitrescu, A., & Zeanah, C. H. (2002). Attachment disturbances in young children: I.

The continuum of caretaking casualty. *Journal of the American Academy of Child and Adolescent Psychiatry, 41*, 972–982.

Smyke, A. T., Koga, S. F. M., Johnson, D. E., Zeanah, C. H., & the BEIP Core Group. (2007). The caregiving context in institution reared and family reared infants and toddlers in Romania. *Journal of Child Psychology and Psychiatry, 48*, 210–218.

Stephenson, P. A., Anghelescu, C., Bobe, N., Ciomartan, T., Gaistenau, M., Fumarel., S., et al. (1993). The causes of children's institutionalization in Romania. *Child: Care, Health, and Development, 20*, 77–88.

Teti, D. M. (1999). Conceptualizations of disorganization in the preschool years: An integration. In J. Solomon & C. George (Eds.), *Attachment disorganization* (pp. 213–242). New York: Guilford.

Tizard, B., & Hodges, J. (1978). The effect of early institutional rearing on the development of eight year old children. *Journal of Child Psychology and Psychiatry, 19*, 99–118.

Tizard, B., & Rees, J. (1974). A comparison of the effects of adoption, restoration to the natural mother, and continued institutionalization on the cognitive development of four-year-old children. *Child Development, 45*, 92–99.

Tizard, B., & Rees, J. (1975). The effect of early institutional rearing on the behavior problems and affectional relationships of four-year-old children. *Journal of Child Psychology and Psychiatry, 16*, 61–73.

Toth, S., Maughan, A., & Manly, J. T. (2002). The relative efficacy of two interventions in altering maltreated preschool children's representational models: Implications for attachment theory. *Development and Psychopathology, 14*, 877–908.

van den Boom, D. C. (1994). The influence of temperament and mothering on attachment and exploration: An experimental manipulation of sensitive responsiveness among lower-class mothers with irritable infants. *Child Development, 65*, 1457–1477.

van IJzendoorn, M. (1995). Adult attachment representations, parental responsiveness, and infant attachment: A meta-analysis on the predictive validity of the Adult Attachment Interview. *Psychological Bulletin, 117*, 387–403.

Vondra, J. I., Hommerding, K. D., & Shaw, D. S. (1999). Stability and change in infant attachment style in a low-income sample. In J. I. Vondra, & D. Barnett (Eds.), Atypical attachment in infancy and early childhood among children at developmental risk. *Monographs of the Society for Research in Child Development*, 64 (3, Serial No. 258), 119–144.

Vorria, P., Papaligoura, Z., Dunn, J., van IJzendoorn, M. H., Steele, H., Kontopoulou, A., et al. (2003). Early experiences and attachment relationships of Greek infants raised in residential group care. *Journal of Child Psychology and Psychiatry, 44*, 1–13.

Wassenaar, D. R. (2006). Commentary: Ethical considerations in international research collaboration: The Bucharest Early Intervention Project. *Infant Mental Health Journal, 27*, 577–580.

Zeanah, C. H., Koga, S. F., Simion, B., Stanescu, A., Tabacaru, C. L., Fox, N. A., et al. (2006). Ethical considerations in international research collaboration: The Bucharest Early Intervention Project. *Infant Mental Health Journal, 27*, 559–576.

Zeanah, C. H., Nelson, C. A., Fox, N. A., Smyke, A. T., Marshall, P., Parker, S., et al. (2003). Effects of institutionalization on brain and behavioral development: The Bucharest Early Intervention Project. *Development and Psychopathology, 15*, 885–907.

Zeanah, C. H., Scheeringa, M. S., Boris, N. W., Heller, S. S., Smyke, A. T., & Trapani, J. (2004). Reactive attachment disorder in maltreated toddlers. *Child Abuse and Neglect, 28*, 877–888.

Zeanah, C. H., Smyke, A. T., Koga, S. F. M., Carlson, E., & The BEIP Core Group (2005). Attachment in institutionalized and non-institutionalized Romanian children. *Child Development, 76*, 1015–1028.

Child Development, January/February 2010, Volume 81, Number 1, Pages 224–236

Neurodevelopmental Effects of Early Deprivation in Postinstitutionalized Children

Seth D. Pollak
University of Wisconsin

Charles A. Nelson
Harvard Medical School/Children's Hospital Boston

Mary F. Schlaak and Barbara J. Roeber
University of Wisconsin

Sandi S. Wewerka, Kristen L. Wiik, Kristin A. Frenn, Michelle M. Loman, and Megan R. Gunnar
University of Minnesota

The neurodevelopmental sequelae of early deprivation were examined by testing ($N = 132$) 8- and 9-year-old children who had endured prolonged versus brief institutionalized rearing or rearing in the natal family. Behavioral tasks included measures that permit inferences about underlying neural circuitry. Children raised in institutionalized settings showed neuropsychological deficits on tests of visual memory and attention, as well as visually mediated learning and inhibitory control. Yet, these children performed at developmentally appropriate levels on similar tests where auditory processing was also involved and on tests assessing executive processes such as rule acquisition and planning. These findings suggest that specific aspects of brain-behavioral circuitry may be particularly vulnerable to postnatal experience.

Over the past decade, increased attention has been devoted to the development of children who have spent some or all of their lives in institutional care (Johnson, 2001). The increase in adoption of institutionalized children has heightened concerns about long-term effects of early deprivation. While the deprivation experienced by children in institutional settings is often impossible to accurately quantify, the environments many of these children endure fall below the quality needed to sustain normal physical and behavioral development. As evidence, institutionalized infants and toddlers lose about 1 month of linear growth for every 2–3 months in institutional care (Johnson, 2001), with behavioral development exhibiting similar dramatic delays (Gunnar, 2001). When institutionalized children are placed in families, marked improvements in physical, social, and cognitive functioning are typically observed, yet many of the children maintain persistent behavioral problems (Ames, 1997; Hodges & Tizard, 1989; Rutter, 1998; Verhulst, Althaus, & Versluis-denBieman, 1990, 1992). The developmental difficulties experienced by many of these children raise questions about the effects of early deprivation including factors such as failure to provide adequate nutrition, medical care, stimulation, and the lack of consistent and supportive caregiving relationships. Although early research emphasized the significance of maternal deprivation, Rutter (1981) rightly noted that many other types of stimulation needed for normal development are also deficient in these environments.

The critical questions that emerge from the plight of these children concern which aspects of inadequate stimulation result in cascading developmental effects, which developmental processes are most affected by inadequate early care, and specification

---

This project was supported by grants from the National Institute of Mental Health to Seth Pollak (R01 MH068858) and Megan Gunnar (R01 MH068857). Support for Charles A. Nelson comes from the National Institutes of Health (NS034458; MH078829) and the Richard David Scott endowment. Infrastructure support was provided by the Waisman Center, University of Wisconsin through the National Institute of Child Health and Human Development (P30-HD03352, M. Seltzer, Director). We acknowledge the research assistance and support provided by Meg Bale. We also greatly appreciate the children and their families whose participation made this research possible.

Correspondence concerning this article should be addressed to Megan R. Gunnar, Institute of Child Development, 51 East River Road, University of Minnesota, Minneapolis, MN 55455. Electronic mail may be sent to gunnar@umn.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8101-0015

about how the transfer to more normative caregiving can foster growth and recovery following institutionalization. Institutionalized children have experienced highly species atypical deprivation; in many countries the institutional conditions are quite poor: Children may be confined to cots, fed gruel through propped up bottles, lack toys or stimulation, and receive very little linguistic stimulation and/or one-to-one interaction with caregivers (Nelson, 2007; Rutter, 1998). Even in institutions where basic physical needs were met, lack of individualized care and attention remain prominent. At adoption, children generally move into middle- to upper-middle-class families, who are generally highly stable and well educated (Hellerstedt et al., 2008). In short, adoption marks a dramatic termination of deprivation, allowing an examination of the impact of early deprivation and neglect on subsequent development. Adoption into a supportive home can provide a profound natural intervention in the life of a child exposed to significant early adversity (van IJzendoorn, Juffer, & Poelhuis, 2005). As a result, internationally adopted children provide a natural experiment on the impact of different degrees and duration of care on subsequent biobehavioral development. Despite occasional significant adjustment problems of the children, there are very few adoption disruptions for families who adopt internationally (e.g., Brumble, 2007). Furthermore, all studies of children adopted or fostered from institutions have shown that, varying with duration of institutionalization, once out of the institution children begin to show remarkable rebounds in physical and cognitive development (Kreppner et al., 2007; Maclean, 2003). Both the capacity of children to rebound after early institutional care and limitations on recovery imposed by longer periods of institutionalization were recently demonstrated experimentally in a study involving random assignment to high-quality foster care for children who began their lives in Romanian institutions (Nelson et al., 2007). Studying postinstitutionalized (PI) children several years after adoption allows examination of long-term impacts of early experience on children's development.

The present study is specifically motivated by the convergence of behavioral studies of PI children and experimental studies of deprivation on the nonhuman primates (see review by Sanchez & Pollak, 2009). In terms of domains where children do not appear to catch up in the immediate years following adoption, Beckett et al. (2007) reported that children adopted after a prolonged period of institutional care from Romania had significantly lower scholastic attainment scores than those adopted early (before 6 months) either within the United Kingdom or from Romania. A separate research team also concluded that although adopted children showed age-expected development in some domains, adoptions after 12 months of age were associated with problems in school achievement (van IJzendoorn & Juffer, 2006). Another recent report indicated that children residing within Romanian orphanages had poorly developed language abilities (Windsor et al., 2007). One recent study focused specifically on language and cognitive outcomes at 6 and 11 years of age with a large sample of institutionally reared Romanian children adopted into UK families. This study noted few negative effects of deprivation if institutionalization ended before the age of 6 months. Even for the children over 18 months, the presence of even very minimal language skills (the child's ability to imitate speech sounds) at the time of arrival was a strong positive prognostic factor for school-aged cognitive outcomes. Importantly, variations in adoptive parent characteristics were unrelated to differences in the adopted children's cognitive outcomes (Croft et al., 2007). Studies of institutionally reared children yield consistent evidence that early deprivation can have long-term consequences for cognitive functioning and school readiness. Yet, most extant research has employed gross measures of functioning such as checklists, questionnaires, school records, and global developmental quotient (DQ) and intelligence quotient (IQ) tests. While such global measures provide some suggestion of which neural systems may have been affected, they are not specific enough to test hypotheses about the development of neural systems. Our goal was to more directly examine brain–behavior processes in systems believed to underlie children's scholastic performance, such as aspects of attention, inhibitory control, working memory, and learning.

These domains are of interest not only because of the persistent difficulties observed in previously neglected children but because nonhuman primate studies suggest that the prefrontal cortex (PFC) and its associated systems are particularly relevant to understanding cognitive functions and may be especially vulnerable to early experience (e.g., Sanchez, Ladd, & Plotsky, 2001). The PFC develops over a protracted period, well into adolescence (Huttenlocker, 1990; Sowell, Thompson, Holmes, Jernigan, & Toga, 1999). While the timing of development in different regions has not been fully determined, there is evidence that neuronal organization in the dorsolateral prefrontal cortex (DL PFC; Brodman areas 9 and 10) changes dramatically until 5–7 years of age (Blinkov & Glezer, 1968) and slows thereafter. A similar protracted period has been observed for the neurotransmitter systems mediating prefrontal activity. Bourgeois (2000) has plotted the overproduction and pruning of synapses in the rhesus monkey against that of the human, taking into consideration differences in age compression (i.e., the rate at which each species develops), and has reported remarkable similarity across species. Collectively, these data, coupled with the postmortem data, point to a protracted period of PFC development. Presumably, this may allow experience-dependent fine-tuning of attention, learning, emotion, and memory systems (e.g., Black, Jones, Nelson, & Greenough, 1998).

Although it is critical to understand whether and how postnatal deprivation and neglect influence brain development in human children, the human neuroscience evidence is sparse. Two studies conducted by Chugani and colleagues reported that PI children from Romania showed significantly decreased metabolism bilaterally in the orbital frontal gyrus, the infralimbic PFC, the medial temporal structures (amygdala and head of hippocampus), the lateral temporal cortex, and the brain stem as compared to normal adults and children with chronic epilepsy. Diffusion tensor imaging also suggested that the PI children had reduced white matter tracts between the anterior temporal and frontal lobes (Chugani et al., 2001; Eluvathingal et al., 2006). These brain imaging findings are consistent with recent work from the Bucharest Early Intervention Project (BEIP; see Zeanah et al., 2003). For example, this group has reported that children reared in institutions show dramatic reductions in brain activity as revealed by the electroencephalogram and the event-related potential (ERP; see Marshall et al., 2004; Parker, Nelson, & the BEIP Core Group, 2005a, 2005b). Importantly, children placed in high-quality foster care before the age of 2 years show improvement in their electroencephalogram (Marshall, Reeb, Fox, Nelson, & Zeanah, 2008); ERP amplitude also improves with placement in foster care, although it is not time- or sensitive-period dependent (Moulson, Fox, Zeanah, & Nelson, 2009).

The present study sought to determine which domains of cognitive development may be particularly affected by institutional neglect and deprivation using children adopted internationally from orphanages or other institutions. These PI children were (a) adopted over age 12 months, (b) spent at least 75% of their lives prior to adoption in institutional care, and (c) resided in their adoptive families for a minimum of 3 years at the time of testing. We compared these children to two different groups. Our first comparison group consisted of children who were adopted before 8 months predominantly from foster care overseas, having little or no institutional care history. This group helps to control for prenatal factors and heritable factors associated with a child becoming orphaned or abandoned. Our second comparison group consists of nonadopted (NA) children who were reared in their families of origin. As noted by Rutter, Dunn, Plomin, and Simonoff (1997), even when using "normed" neuropsychological evaluations, and especially when using nonstandardized tests, it is critical to include NA children who have grown up in families of similar economic and educational histories to the families of adopted children. Furthermore, for domains that yielded differences between the PI children and both other groups, we examined whether duration of institutional care correlated with the children's performance. Thus, the strategy employed here is to identify as "institutional deprivation effects" those cognitive functions for which the PI group differs from the two comparison groups and, further, whether the duration of institutional care is associated with the outcome within the PI group. Here, we focus on basic cognitive and learning processes that might underlie school-based learning problems.

Method

Participants

The participants were 132 children ages 8 years 0 months to 9 years 11 months. Three groups were examined (see Table 1 for details). A PI group whose selection criteria were adopted at 12 months of age or older (range = 12–78 months, *M* = 23.4 months, *SD* = 12.9) having spent at least 75% of their preadoption lives in institutional care. Over 50% of this group had no experience other than institutional care prior to adoption (*M* = 22.1 months in institutional care, *SD* = 12.4). Children in the PI group were adopted from Asia (*n* = 19), Latin America (*n* = 1), Russia and Eastern Europe (*n* = 27), and Africa (*n* = 1). An early adopted predominantly from foster care (EA) group whose selection criteria were adoption at 8 months or earlier (range = 2–8 months, *M* = 5.2 months, *SD* = 1.7 months) having spent 2 months or less in institutional care. Nearly 83% of this group had spent all of their preadoption lives in foster care overseas.

Table 1
Sample Characteristics

	Postinstitutionalized (*N* = 48)	Early adopted (*N* = 40)	Control (*N* = 44)
Sex (% female)	50	47	55
Age	8 years 4 months (0.5 months)	8 years 4 months (0.5 months)	8 years 5 months (0.5 months)
IQ[a]	105.9 (15.8)	113.7 (16.5)	124.7 (15.7)
Years of parent education	16.3 (2.2)	16.8 (1.6)	16.4 (2.0)
Median family income	75–110K	75–110K	50–75K

[a]Estimated based on Vocabulary and Block Design subtests of the WIS Wechsler Intelligence Scale for Children, 3rd ed. (C–III), or the Leiter International Performance Scale, Revised (Leiter–R) for participants who scored below the normal range on the WISC–III.

Those with institutional experience had been adopted before 3 months of age. Children in the EA group were adopted from Asia (67%) and Latin America (33%). The EA group permitted comparison with children who had experienced loss of birth families and adoption into another culture. Finally, the third group was children born and raised in their birth families in the United States (NA). The children were recruited and tested at two sites, the University of Wisconsin ($n = 58$) and the University of Minnesota ($n = 74$). As shown in Table 1, there were no differences between the three groups in terms of the numbers of boys versus girls, child's age, or parent education. As expected, children in the PI group had lower IQ scores than those children in the two (EA and NA) comparison groups, $F(2, 131) = 16.13$, $p < .001$. On average, adoptive parents had higher family incomes than those from control families, $F(2, 131) = 3.99$, $p = .02$.

Recruitment and Screening

The internationally adopted children were drawn from the Minnesota and the Wisconsin International Adoption Project Registries—registries of families created through international adoption that expressed interest in being contacted about research participation. The NA children were recruited in Wisconsin through fliers and advertisements and in Minnesota from the Institute of Child Development Participant Pool, a registry of children whose parents indicated interest in being contacted about research opportunities in response to a mailing soon after their children were born.

Within 6 months of the present study, the children were all screened to determine if they met the study's group assignment (described earlier) and exclusion criteria. Criteria for group assignment were determined through phone interview and parent questionnaire. Children then participated in an extensive developmental profile, parts of which determined exclusion criteria. Children were excluded if their IQs were below the normal range (< 78), parents reported congenital abnormalities (e.g., Down syndrome or cerebral palsy), and failure on the fetal alcohol syndrome (FAS) screener. We screened for IQ using the Vocabulary and Block Design subtests of the Wechsler Intelligence Scale for Children, 3rd ed. (WISC–III; Weschler, 1998). If children scored below 1 *SD* on either WISC–III subtests, they were subsequently administered the Leiter International Performance Scale, Revised (Roid & Miller, 1997), a nonverbal IQ assessment, to avoid excluding children with normal IQs whose performance might have been affected by the English language demands of the WISC–III. Children were required to pass either the WISC or the Leiter to be included in this study; if they did not reach criterion on both the WISC and Leiter, then they were excluded. Eight children were assessed with the Leiter and of those, 4 (all PI) were excluded because of IQs in the study exclusion range.

To screen for possible FAS, we photographed front and side views of children's faces and analyzed these images using FAS Facial Photographic Analysis Software (Astley, 2003). Children scoring in the moderate range or higher on this screener were excluded and images where coders were uncertain were also reviewed by a pediatric dysmorphologist. Reliability among coders was 100%. Parents of children who failed the FAS screener were contacted and referred for further evaluation. Four children (2 PI, 2 EA) were excluded because of possible FAS. Note that this exclusion criterion only eliminated children with significant facial dysmorphology due to prenatal alcohol exposure. Children with exposures that did not affect facial morphology would not be identified by this method.

In addition to these assessments, children were also administered the Paragraph Comprehension subtest from the Comprehensive Assessment of Spoken Language battery (CASL). The tasks employed in this study were selected because they required only rudimentary language comprehension. Nonetheless, children scoring 2 *SD* below the published norms of this tests were excluded to reduce the possibility that language problems might impair their performance. None of the PI or EA children needed to be excluded because of low CASL scores.

Four PI (8%), 2 EA (3%), and 1 NA (2%) were on stimulant medications for attention problems; these children were tested when parents agreed to withhold medication (e.g., weekends). Across all of the above-mentioned criteria, 1 NA, 4 EA, and 7 PI children were screened but not included in our final sample of 132.

For descriptive purposes, because low birth weight is associated with poorer cognitive functioning, as part of the screening examination parents provided information about the children's birth weight. Parents were asked to check baby books and other records for this information. Parents who were not able to find these records were asked to sign a medical release so that the child's pediatrician could be contacted for this information. Birth weight data were obtained from 97% of the NA, 85% of the EA, and 41% of the PI children. Notably, in cases in which the PI children were abandonned (e.g., typical for Chinese PI children), this information was unknown. Analysis of the birth weight data yielded means in grams for the PI, EA and NA children, respectively, of 2,893 ($SD = 1{,}134$), 2,982 ($SD = 524$), and 3,314 ($SD = 699$), $F(2, 94) = 5.9$, $p < .01$. Hockbergen post hoc tests indicated that both PI and EA groups were lower birth weight, on average, than NA children, although PI and EA birth weight did not differ. Because of the large number of EA children from Asia, we also compared EA and PI children only from that area of the world. The results failed to yield any difference between Asian EA and PI children, $t < .01$, *df* for unequal *ns* = 5.3, *ns*.

Procedure

Neuropsychological functioning was assessed through two well-validated test batteries. One of these, the Cambridge Neuropsychological Test Automated Battery (CANTAB; Cambridge Cognition, Cambridge, United Kingdom), is a computerized series of neuropsychological tests that cover a wide range of cognitive domains. The second battery is the NEPSY Developmental Neuropsychological Assessment (Korkman, Kirk, & Kemp, 1998), a battery developed specifically for use with children. The CANTAB offers several strengths: It has been used extensively with children (Luciana & Nelson, 2002), it has proved sensitive in discriminating various clinical populations from typically developing children, the individual tasks have been studied extensively to confirm their neural correlates (Joyce & Robbins, 1991; Owen, Downes, Sahakian, Polkey, & Robbins, 1990; Owen, Iddon, Summers, & Robbins, 1997; Owen, Morris, Sahakian, Polkey, & Robbins, 1996; Owen, Roberts, Polkey, Sahakian, & Robbins, 1991; Owen, Sahakian, Semple, Polkey, & Robbins, 1995; Saghal et al., 1991; Sahakian et al., 1988), it is computerized for standardized administration, the stimuli cannot be verbalized, and the subtasks require nonverbal responses; thus, performance is not confounded with subjects' verbal skills.

Neuropsychological test. Memory functions were evaluated with three different tests. The first, spatial working memory, tests the child's ability to retain spatial information and to manipulate remembered items in working memory. It is a self-ordered task, which also assesses heuristic strategy. The test begins with a number of colored squares displayed on a computer monitor. By touching the boxes and using a process of elimination, the child should find one blue "token" in each of a number of boxes and use them to fill up an empty column on the right-hand side of the screen. The number of boxes is gradually increased until it is necessary to search a total of eight boxes. The child's score is based upon errors (touching boxes that have already been found to be empty, and revisiting boxes that have already been found to contain a token) as well as response latency, which is age adjusted. Memory for faces assesses the ability to recognize faces after a single exposure. Children are presented 16 pictures of children one at a time for 5 s each, during which children are instructed to remember the faces and identify the sex of each child. Next, the children are presented with pictures of 3 children, 1 of which they previously saw, and they are then asked to identify the picture they saw previously. Thirty min later, this procedure is repeated. The score on this test is based on the sum of correct responses in immediate and delay tasks. Next, we administered a narrative memory test that assesses the ability to retell a story under both free and cued recall conditions. A story is told to the child. Following the story, children are asked to tell the examiner the same story. After the child has recalled as much as they can, the examiner cues details the child did not mention.

Learning processes were evaluated using the paired associates learning test, which assesses visual memory and new learning. Boxes are displayed on the screen and are opened in a randomized order. One or more of them will contain a pattern. The patterns are then displayed in the middle of the screen, one at a time, and the subject must touch the box where the pattern was originally located. If the subject makes an error, the patterns are re-presented to remind the subject of their locations. The difficulty level increases through the test.

Attentional processes were evaluated using five different tasks. The auditory attention and response set test assesses children's ability to maintain selective auditory attention, as well as the child's ability to maintain a complex mental set. In the first part of the task, the child learns to manipulate a red item whenever they hear the word *red*. In the second part of the task, the child then must shift set and respond to red stimuli when hearing the word *yellow*. Match to sample visual search is a matching test, with a speed–accuracy trade-off. Children are shown a complex visual pattern in the middle of a computer screen, and then, after a brief delay, a varying number of similar patterns are shown in a circle of boxes around the edge of the screen. Only one of these boxes matches the pattern in the center of the screen, and the subject must indicate which it is by touching it. Reaction time is measured on the basis of the release of the press pad, which allows for its more accurate measurement. Rapid visual information processing is a test of visual sustained attention. A white box appears in the center of the computer screen, inside which digits appear in a pseudo-random order, at the rate of 100 digits per minute. Children are requested to detect target sequences of digits and to register responses using a press pad. Outcome measures are calculated using signal detection theory. The visual attention test assesses the speed and accuracy with which a child can scan an array and locate targets. Finally, the knock and tap test measures the child's ability to inhibit immediate impulses evoked by visual stimuli that conflict with a verbal direction. The child learns a pattern of motor responses and then must maintain that cognitive set and inhibit the impulse to imitate the examiner's action.

The working memory and knock and tap tasks also involve components of executive functions. We examined two additional aspects of executive function, rule acquisition and manipulation through reversal and planning, here labeled *executive processes*. The intra–extra dimensional shift set (ID/ED) test was used to measure rule acquisition and manipulation through reversal. Two artificial dimensions are used in the test: color-filled shapes and white lines. Simple stimuli are made up of just one of these dimensions, whereas compound stimuli are made up of both, namely, white lines overlying color-filled shapes. Children begin by seeing two simple color-filled shapes and must learn which one is correct by touching it. Feedback teaches the child which stimulus is correct, and after six correct responses, the stimuli and/or rules are changed. These shifts are initially intradimensional (e.g., color-filled shapes remain the only relevant dimension), then later extradimensional (white lines become the only relevant dimension). Children progress through the test by satisfying a set criterion of learning at each stage. In addition, Stockings of Cambridge is a version of the Tower of London spatial planning test, which gives a measure of frontal lobe function (Baker et al., 1996; Owen et al., 1990; Robbins, 1996). Children are shown two displays containing three colored balls. The displays are presented in such a way that they can easily be perceived as stacks of colored balls held in stockings or socks suspended from a beam. The child must use the balls in the lower display to copy the pattern shown in the upper display. The balls may be moved one at a time by touching the required ball, then touching the position to which it should be moved. The time taken to complete the pattern and the number of moves required are taken as measures of the child's planning ability.

Results

Our first set of analyses sought to determine which domains of functioning were either minimally affected by early deprivation or may reflect developmental catch-up following family rearing. Although we could not distinguish between these explanations, we could determine tasks in which our three groups of children performed differently. Omnibus tests revealed group differences in memory functioning, $F_{\text{mult}}(6, 252) = 3.31$, $p = .004$; on tests of attentional functioning, $F_{\text{mult}}(10, 224) = 2.72$, $p = .004$; and on the test of learning, $F(2, 131) = 12.47$, $p = .001$. However, all three groups of children performed similarly on tests of executive processes, $F_{\text{mult}}(4, 254) = 1.20$, *ns*.

Next, we explored those domains (memory, attention, and learning) where multivariate tests suggested group differences to determine which specific tasks discriminated between groups of children. Table 2 shows the descriptive information for

Table 2
Descriptive Data on Domains Exhibiting Significant Group Differences

		Groups			
	Postinstitutionalized	early adopted	Control	Post hoc comparisons	Duration institutional care
Memory tasks					
Spatial working memory[a] (% errors)	52.7 (18.4)	40.9 (19.5)	39.0 (15.7)	PI < Con EA = Con	−.18
Memory for faces[b] (scaled score)	11.1 (3.1)	12.2 (2.8)	12.6 (3.3)	PI < Con EA = Con	−.11
Auditory narrative memory[b] (scaled score)	9.7 (4.1)	10.8 (3.3)	11.4 (3.2)	No significant difference	−.31*
Attention tasks					
Match to sample[a] (% correct)	95.3 (7.3)	96.2 (4.6)	97.4 (3.4)	No significant difference	−.16
Auditory attention[b] (scaled score)	9.2 (1.3)	9.6 (1.4)	9.7 (1.0)	No significant difference	−.06
Rapid visual processing[a] (% correct)	10.5 (6.1)	9.6 (3.9)	8.9 (4.6)	No significant difference	−.04
Visual attention[b] (scaled score)	10.1 (2.3)	12.2 (2.7)	11.7 (2.1)	PI < Con EA = Con	−.33*
Knock and tap[b] (raw score)	28.00 (2.93)	28.69 (1.81)	29.10 (1.05)	PI < Con EA = Con	−.35*
Executive control tasks					
ID/ED[a] (total errors)	41.2 (19.3)	38.3 (23.7)	33.7 (17.6)	No significant difference	.09
Stockings of Cambridge[a] (problems solved)	6.8 (1.8)	7.3 (1.7)	7.0 (1.5)	No significant difference	−.14
Learning task					
Paired associates learning[a] (total errors)	14.6 (13.8)	6.9 (5.9)	6.2 (5.2)	PI < Con EA = Con	−.46**

Note. EA = early adopted; ID/ED = intra–extra dimensional shift set; PI = postinstitutionalized.
[a] CANTAB subtest. [b] NEPSY subtest.
*$p < .05$. **$p < .01$.

these tasks, the pattern of post hoc comparison between groups, and the correlations of scores with duration of institutional care. We included sex in our analyses, but in no instance did we find an interaction of sex and group. Where main effects of sex were noted, girls performed better than boys; however, as sex differences did not modify the main effects of group, we do not discuss them further. Information on sex differences is available upon request.

Within the memory domain, the groups performed differently on the spatial working memory test, $F(2, 128) = 7.96$, $p = .001$. Hochberg post hoc tests revealed that PI children performed more poorly than both EA ($p = .008$) and controls ($p = .001$), but that EA children performed similarly to controls (*ns*). The memory for faces test resulted in a marginal group difference, $F(2, 128) = 2.95$, $p = .056$. On this test, PI children scored lower than NA children ($p = .057$), but the EA and NA groups performed similarly. The groups did not differ on auditory memory ability.

In terms of tests of learning, children's performance on the paired associates learning test differed, $F(2, 131) = 12.47$, $p = .001$. PI children performed lower than EA ($p = .003$) and NA ($p < .001$), but EA and NA groups performed similarly (*ns*).

Within the attention domain, groups performed similarly on tests of auditory attention $F(2, 116) = 2.30$, *ns*; match to sample $F(2, 116) = 1.75$, *ns*; and rapid visual processing $F(2, 116) = 1.03$, *ns*. However, the groups differed on the test of visual attention, $F(2, 116) = 8.96$, $p = .001$. Here, PI children performed more poorly than EA ($p = .001$) and controls ($p = .006$), but EA children performed similarly to controls ($p = ns$). There was also a marginal difference on the knock and tap test, $F(2, 116) = 2.92$, $p = .06$. Here, PI children showed more impulsivity than controls ($p = .05$), but there were no other differences between groups.

Discussion

The primary goal of this study was to determine which cognitive processes differentiated children who experienced early institutional deprivation and neglect. PI internationally adopted children differed from both groups of comparison children on tests of visual memory and attention, as well as visually mediated learning and inhibitory control. Yet, these children performed at developmentally appropriate levels on similar tests where auditory processing was also involved. These children also performed well on tests of executive processes involving rule acquisition, manipulation, and planning.

This study was not designed to show precisely which neural circuits are affected by early experience, but the tasks that PI children had difficulty with do provide a useful guide for specifying neuroanatomical substrates that can be more directly examined in future studies. Next, we briefly highlight areas of brain circuitry associated with these tasks. Performance on the memory for faces task reflects an interaction between PFC and stimulus-specific visual cortical association areas that mediate visual working memory. The neurophysiological models of visual working memory developed in the nonhuman primate also holds for humans, based upon event-related functional magnetic resonance imaging (ER-fMRI) studies of facial delay recognition task similar to the one used in the present study (Druzgal & D'Esposito, 2003). Memory tasks such as the one used in the present study involve both visual memory and visual perception. But there is not yet agreement in the field about the neural bases of these processes. Some data suggest that visual memory and visual perception are associated with common neural substrates (Slotnick, 2004), whereas other data suggest that right medial temporal lobe structures are critically involved in retention, but not in the perception, of new faces (Crane & Milner, 2002).

Spatial working memory is a test of the child's ability to retain spatial information and to manipulate remembered items in working memory and also assesses heuristic strategy. The task is considered a sensitive measure of frontal lobe dysfunction based upon studies of nonhuman primates and patients with frontal lobe damage (Owen, Morris, et al., 1996; Owen et al., 1990; Owen et al., 1995; Passingham, 1995). Owen, Evans, and Petrides (1996) used positron emission tomography with magnetic resonance imaging to demonstrate the existence, within the human brain, of two functionally distinct subdivisions of the lateral frontal cortex, which subserve different aspects of spatial working memory. When the spatial working memory task required the organization and execution of a sequence of spatial moves retained in working memory, significant changes in blood flow were observed in ventrolateral frontal cortex (area 47) bilaterally. In contrast, when the task required active monitoring and manipulation of spatial information within working memory, additional activation foci were observed in mid-dorsolateral frontal cortex (areas 46 and 9). Both of these processing stages were required to successfully complete a spatial working memory task (Owen, Evans, et al., 1996). Another study also linked poor performance on the CANTAB spatial working memory task in children to reduced dorsolateral PFC using fMRI (Luna et al., 2002).

It is difficult to link performance on tests of visual attention to distinct neural systems. A recent study used ER-fMRI to measure brain activity as subjects oriented visual attention to an upcoming target. This activation was lateralized to the left hemisphere and reflected a widely distributed network that included (a) structures in parietal and temporal cortices and thalamus usually associated with selective attention; (b) ventral-stream object processing structures in occipital, inferior-temporal, and parahippocampal cortex; and (c) structures in medial- and dorsolateral prefrontal cortex associated with cognitive control (Arrington, Carr, Mayer, & Rao, 2000). In visual attention tasks, subjects will undoubtedly make errors and need to reorient to targets. This same study revealed that brain areas specific to attentional reorientation were right-lateralized and included posterior temporal and inferior parietal regions, as well as prefrontal regions that likely subserve control processes related to inhibition of inappropriate responding (Arrington et al., 2000). In sum, selective visual attention results from distributed brain activity that includes maintenance of the selected object's representation accompanied by suppression of response to ignored objects (Duncan, 1993; Farah, 1990; Phaf, Van der Heijden, & Hudson, 1990), with control of these processes mediated by the medial DL PFC.

Paired associates learning is a stringent test for visual episodic memory and associative learning (Sahakian & Owen, 1992). To perform well, children had to learn the locations of a progressively increasing number of abstract stimuli. Although such a complex behavioral task undoubtedly draws upon widely distributed neural systems, ER-fMRI suggests that successful performance on this type of task relies heavily upon medial temporal lobe connections to the frontal lobe (Aizenstein et al., 2000). Similarly, the knock and tap test cannot be mapped onto discrete circuitry. The test measures the child's ability to inhibit immediate impulses evoked by visual stimuli that conflict with a verbal direction. The child learns a pattern of motor responses and then must maintain that cognitive set and inhibit the impulse to imitate the examiner's action.

Taken together, the present data suggest delayed maturation of select aspects of frontal circuitry, and perhaps reduced functional connectivity of frontal cortex with other neocortical and subcortical regions, plays a key role in scholastic difficulties among children who experienced early institutionalized deprivation and neglect. Indeed, regions of the PFC have long been associated with cognitive processes similar to the ones assessed in this study (Fletcher & Henson, 2001; Owen, Morris, et al., 1996). Studies in nonhuman primates, lesion studies, as well as functional neuroimaging studies in humans, have documented that DLPFC is crucial for maintaining and manipulating information in ways assessed by the tasks used here (Carlson et al., 1998; Fuster, 2000; Goldman-Rakic, 1988; McCarthy et al., 1994; Owen, Evans, et al., 1996; Pierrot-Deseilligny, Rivaud, Gaymard, Muri, & Vermersch, 1995; Sweeney et al., 1996).

The distinction between the performance of PI and EA children suggests (but does not prove) that the ontogenesis of the PFC also includes an extensive postnatal interval. Such an interpretation is consistent with animal studies that reveal, for example, that the thickness of the PFC of the rat brain is not maximal until postnatal day 20, which marks the beginning of the postweaning period (e.g., roughly equivalent to adolescence in humans; Vincent, Khan, & Benes, 1995). Indeed, PI children are often noted to have problems in attention regulation and emotional control, functions presumably influenced by the development of the PFC and its distributed systems (see review by Gunnar, 2001; Shallice et al., 2002). Consistent with this view, Sanchez, Hearn, Do, Rilling, and Herndon (1998) studied rhesus monkeys that were socially deprived between 2 and 12 months of age. These monkeys exhibited cognitive deficits that had also been noted in earlier studies (e.g., Harlow, Harlow, & Suomi, 1971). Sanchez et al.'s MRI studies revealed that the animals' performance on executive function tasks was correlated with decreased neuronal development of prefrontal, medial temporal, and amygdala substrates. Two years later, these monkeys exhibited increased abnormalities in the PFC that were related to the monkeys' performance on learning and working memory tasks (Sanchez et al., 2001). In a separate study, Mathew et al. (2003) reported neuropathological alterations in the PFC of adult macaques with early adverse experience.

There are a number of features of the present study that are important to consider in interpreting these data. First, as described earlier, we excluded children with clear developmental delays or indications of fetal alcohol exposure, which would have artificially exaggerated group performance deficits. By avoiding a situation where a subset of PI children may be experiencing global cognitive delays, we were also able to detect specific patterns of processing deficits. In general, executive function tasks are sensitive to perturbations in children's lives, but effects on these tasks are not specific to certain situations or conditions. If we observed broad or diffuse cognitive deficits in the PI sample, it would not be possible to rule out general attentional, IQ, motivational, motor, or emotional problems as driving children's performance. Yet, in the present study we observed that on a majority of tasks, PI children displayed performance equivalent to their peers, which permits stronger interpretation of those specific areas where deficits emerged. Second, this sample of children represents families who volunteered to participate in research; it may be the case that parents who believed their children were doing particularly well or poorly were more likely to participate. Yet, such sampling issues—if they exist—do not readily explain the pattern of results observed in this study. In addition, our unusual sampling procedures (see Hellerstedt et al., 2008) probably resulted in a group of children that are both diverse and representative.

The study of children reared in atypical situations requires opportunistic approaches that are also fraught with interpretive challenges. Our requirement that the EA group be adopted before 8 months of age and that these children had to spend ≤2 months in institutional care meant that none of the EA children were drawn from Eastern Europe, where adoption processes take more time. Most of the EA children spent most of their preadoptive lives in foster care overseas, an option that did not exist in Eastern Europe between 1990 and 1998, when the children that we tested were adopted. It is likely that prenatal conditions were less optimal for children who end up being placed in orphanage or institutional care overseas than for children born and raised in their families of origin in the United States. In part, the inclusion of a group of children adopted internationally with little or no institutional experience partly addressed this problem. We obtained birth weight data on the majority of both EA and NA children, and found, as might be expected, that the EA children were lower in birth weight than the NA children. Although we were only able to obtain birth weight data on 41% of the PI children, there was no evidence that birth weight differentiated PI from EA children. While we must be cautious in interpretation of these data because we lack evidence on the reliability of the information and are missing substantial data for the PI children, the pattern of data suggests that EA and PI children tended to experience prenatal environments that were impoverished. This tends to clarify, but not resolve, the issue of whether pre- versus postnatal environmental impacts influenced the outcomes we examined.

In terms of the tasks administered to children, it is critical that measurement involves a sufficient level of task difficulty relative to the ability level of participants. In this regard, the specific tasks used in this study are particularly useful. For example, the test of spatial working memory ability and paired associates learning systematically varies the working memory load, which increases the amount of information that needs to be remembered and the number of trials over which it needs to be maintained. The children's pattern of performance across the tasks that we administered did not reveal clear performance deficits on auditory memory and attentional tasks. One possibility is that the deficits observed here are a function of how much information needs to be kept active in working memory over time, and how precisely that information needs to be encoded. At the same time, firm conclusions about neural activity cannot be made based solely on behavioral data. However, these data do underscore the need to employ specific assessments that have been used in imaging and lesion studies and that may allow more fine-grained analysis of the impact of institutional neglect on neurobehavioral development.

Why did the PI children perform better on tasks that relied primarily on auditory, as compared with visual, information? One possibility is that visual development is more vulnerable to postnatal influences. The auditory system starts functioning during the last trimester of gestation (Birnholz & Benaceraff, 1983), whereas the visual system does not start functioning until after birth. Thus, both visual and auditory *functional* development reflect experience-dependent processes; the difference is that auditory experience starts before visual experience, when the brain is at a different point in its development (cf. Kellman & Arterberry, 2006; Saffran, Werker, & Werner, 2006). Sloutsky and colleagues (Robinson & Sloutsky, 2004; Sloutsky & Napolitano, 2003) have demonstrated that young children, unlike adults, exhibit auditory dominance. This may reflect earlier maturation of the auditory system relative to the visual system. It may also be the case that attentional processes more easily engage some stimulus properties. For example, auditory stimuli are serial, transient events that must be perceived quickly whereas visual stimuli are presented episodically for longer periods of time. Thus, attentional systems may allocate more resources to transient relative to stable stimuli. Similarly, Posner, Nissen, and Klein (1976) proposed that attention to visual stimuli must be learned, whereas attention to auditory stimuli is more automatic. Also consistent with the argument that postnatal experience is more likely to influence visual processing, Maurer, Lewis, Brent, and Levin (1999) have demonstrated that the neural circuitry responsible for adults' face expertise is not prespecified but requires early visual experience (see also Nelson, 2001; Pascalis et al., 2005). Because infants have poor visual acuity, their cortices are only exposed to low spatial frequency. Thus, early exposure to visual information sets up the neural architecture for more complex visual processing. These studies have demonstrated that when visual input is delayed by as little as 2 months, permanent visual deficits result (Le Grand, Mondloch, Mauer, & Brent, 1999, 2003; Maurer et al., 1999).

Models of the role of experience in neural development, and the mounting information on molecular processes in neural plasticity, indicate that neural activities (i.e., activity-dependent processes) are critical to brain development (see Fox, Levitt, & Nelson, 2010). This implies that in addition to the stimuli available in the environment, active engagement of the environment may be essential in order for some aspects of cognitive development to occur. While these opportunities abound in typical human rearing environments, institutionalized childrearing may restrict the kinds of dynamic experiences and input necessary for some aspects of neurobehavioral development. It is hoped that the more we understand about the specific aspects of neurodevelopment impacted by early deprivation, the more we can focus on identifying targeted intervention and training experiences that would optimize these children's outcomes.

References

Aizenstein, H. J., MacDonald, A. W., Stenger, V. A., Nebes, R. D., Larson, J. K., Ursu, S., et al. (2000). Complementary category learning systems identified using event-related functional MRI. *Journal of Cognitive Neuroscience*, 12, 977–987.

Ames, E. W. (1997). *The development of Romanian orphanage children adopted to Canada* (Final report to the national Welfare Grants Program: Human Resources Development Canada). Burnaby, Canada: Simon Fraser University.

Arrington, C. M., Carr, T. H., Mayer, A. R., & Rao, S. M. (2000). Neural mechanisms of visual attention:

Object-based selection of a region in space. *Journal of Cognitive Neuroscience, 12*, 106–117.

Astley, S. (2003). *FAS Facial Photographic Analysis Software. FAS diagnostic & prevention network*. Seattle: University of Washington.

Baker, S. C., Rogers, R. D., Owen, A. M., Frith, C. D., Dolan, R. J., Frackowiak, R. S. J., et al. (1996). Neural systems engaged by planning: A PET study of the Tower of London task. *Neuropsychologia, 34*, 515–526.

Beckett, C., Maughan, B., Rutter, M., Castle, J., Colvert, E., Groothues, C., et al. (2007). Scholastic attainment following severe early institutional deprivation: A study of children adopted from Romania. *Journal of Abnormal Child Psychology, 35*, 1063–1073.

Birnholz, J. C., & Benacerraf, B. R. (1983). The development of human fetal hearing. *Science, 222*(4623), 516.

Black, J. E., Jones, T. A., Nelson, C. A., & Greenough, W. T. (1998). Neuronal plasticity and the developing brain. In N. E. Alessi, J. T. Coyle, S. I. Harrison, & S. Eth (Eds.), *Handbook of child and adolescent psychiatric science and treatment* (pp. 31–53). New York: Wiley.

Blinkov, S. M., & Glezer, I. I. (1968). *The human brain in figures and tables: A quantitative handbook*. New York: Plenum Press.

Bourgeois, J. A. (2000). Visual hallucinations and psychotic symptoms during treatment with selective serotonin reuptake inhibitors: Is the sigma receptor involved? *Journal of Clinical Psychopharmacology, 20*, 580.

Brumble, K. (2007). Intercountry transracial special needs adoptees: Today's teenagers and young adults. How have they fared? The parent's perspective. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 68*, 1404.

Carlson, S., Martinkauppi, S., Rama, P., Salli, E., Korvenoja, A., & Aronen, H. J. (1998). Distribution of cortical activation during visuospatial n-back tasks as revealed by functional magnetic resonance imaging. *Cerebral Cortex, 8*, 743–752.

Chugani, H. T., Behen, M. E., Muzik, O., Juhasz, C., Nagy, F., & Chugani, D. C. (2001). Local brain functional activity following early deprivation: A study of post-institutionalized Romanian orphans. *Neuroimage, 14*, 1290–1301.

Crane, J., & Milner, B. (2002). Do I know you? Face perception and memory in patients with selective amygdalo-hippocampectomy. *Neuropsychologia, 40*, 530–538.

Croft, C., Beckett, C., Rutter, M., Castle, J., Colvert, E., Groothues, C., et al. (2007). Early adolescent outcomes of institutionally-deprived and non-deprived adoptees. II: Language as a protective factor and a vulnerable outcome. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 48*, 31–44.

Druzgal, T. J., & D'Esposito, M. (2003). Dissecting contributions of prefrontal cortex and fusiform face area to face working memory. *Journal of Cognitive Neuroscience, 15*, 771–784.

Duncan, J. (1993). Coordination of what and where in visual attention. *Perception, 22*, 1261–1270.

Eluvathingal, T. J., Chugani, H. T., Behen, M. E., Juhász, C., Muzik, O., Maqbool, M., et al. (2006). Abnormal brain connectivity in children after early severe socioemotional deprivation: A diffusion tensor imaging study. *Pediatrics, 117*, 2093–2100.

Farah, M. J. (1990). *Visual agnosia: Disorders of object recognition and what they tell us about normal vision*. Cambridge, MA: MIT Press.

Fletcher, P. C., & Henson, R. N. (2001). Frontal lobes and human memory: Insights from functional neuroimaging. *Brain: A Journal of Neurology, 124*, 849–881.

Fox, S. E., Levitt, P., & Nelson, C. A. (2010). How the timing and quality of early experiences influence the development of brain architecture. *Child Development, 81*, 28–40.

Fuster, J. M. (2000). Prefrontal neurons in networks of executive memory. *Brain Research Bulletin, 52*, 331–336.

Goldman-Rakic, P. S. (1988). Topography of cognition: Parallel distributed networks in primate association cortex. *Annual Review of Neuroscience, 11*, 137–156.

Gunnar, M. (2001). Effects of early deprivation: Findings from orphanage-reared children. In C. A. Nelson & M. Luciana (Ed.), *Handbook of cognitive neuroscience* (pp. 617–629). Cambridge, MA: MIT Press.

Harlow, H. F., Harlow, M. K., & Suomi, S. J. (1971). From thought to therapy: Lessons from a primate laboratory. *American Scientist, 59*, 538–549.

Hellerstedt, W. L., Madsen, N. J., Gunnar, M., Grotevant, H. D., Lee, R. M., & Johnson, D. E. (2008). The international adoption project: Population-based surveillance of Minnesota parents who adopted children internationally. *Maternal and Child Health Journal, 12*, 162–171.

Hodges, J., & Tizard, B. (1989). IQ and behavioural adjustment of ex-institutional adolescents. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 30*, 53–75.

Huttenlocker, P. R. (1990). Morphometric study of human cerebral cortex development. *Neuropsychologia, 28*, 517–527.

Johnson, D. E. (2001). The impact of orphanage rearing on growth and development. In C. A. Nelson (Ed.), *The effects of adversity on neurobehavioral development: Minnesota symposia on child psychology* (Vol. 31, pp. 113–162). Mahwah, NJ: Erlbaum.

Joyce, E. M., & Robbins, T. W. (1991). Frontal lobe function in Korsakoff and non-Korsakoff alcoholics: Planning and spatial working memory. *Neuropsychologia, 29*, 709–723.

Kellman, P. J., & Arterberry, M. E. (2006). Infant visual perception. In D. Kuhn, R. S. Siegler, W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (6th ed., pp. 109–160). Hoboken, NJ: Wiley.

Korkman, M., Kirk, U., & Kemp, S. (1998). *NEPSY: A developmental neuropsychological assessment manual*. San Antonio, TX: Psychological Corporation.

Kreppner, J. M., Rutter, M., Beckett, C., Castle, J., Colvert, E., Groothues, C., et al. (2007). Normality and impairment following profound early institutional deprivation: A longitudinal follow-up into early adolescence. *Developmental Psychology, 43*, 931–946.

Le Grand, R., Mondloch, C. J., Mauer, D., and Brent, H. P. (1999). Early visual experience and face processing. *Vision Research, 39*, 3480.

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2003). Expert face processing requires visual input to the right hemisphere during infancy. *Nature Neuroscience, 6*, 1108–1112.

Luciana, M., & Nelson, C. A. (1988). The functional emergence of prefrontally-guided working memory systems in four to eight-year-old children. *Neuropsychologia, 36*, 273–293.

Luciana, M., & Nelson, C. A. (2002). Assessment of neuropsychological function through use of the Cambridge Neuropsychological Testing Automated Battery: Performance in 4- to 12-year-old children. *Developmental Neuropsychology, 22*, 595–624.

Luciana, M., Sullivan, J., & Nelson, C. A. (2001). Individual differences in phenylalanine levels moderate performance on tests of executive function in adolescents treated early and continuously for PKU. *Child Development, 72*, 1637–1652.

Luna, B., Minshew, N. J., Garver, K., Lazar, N. A., Thulborn, K. R., Eddy, W. F., et al. (2002). Neocortical system abnormalities in autism: An fMRI study of spatial working memory. *Neurology, 59*, 834–840.

Maclean, K. (2003). The impact of institutionalization on child development. *Developmental Psychopathology, 15*, 853–884.

Marshall, P.J., Fox, N.A., & the BEIP Core Group. (2004). A comparison of the electroencephalogram between institutionalized and community children in Romania. *Journal of Cognitive Neuroscience, 16*, 1327–1338.

Marshall, P., Reeb, B. C., Fox, N. A., & the BEIP Core Group. (2008). Effects of early intervention on EEG power and coherence in previously institutionalized children in Romania. *Development and Psychopathology, 20*, 861–880.

Mathew, S. J., Shungu, D. C., Mao, X., Smith, E. L., Perera, G. M., Kegeles, L. S., et al. (2003). A magnetic resonance spectroscopic imaging study of adult nonhuman primates exposed to early-life stressors. *Biological Psychiatry, 54*, 727–735.

Maurer, D., Lewis, T. L., Brent, H. P., & Levin, A. V. (1999). Rapid improvement in the acuity of infants after visual input. *Science, 286* (5437), 108–110.

McCarthy, G., Blamire, A. M., Puce, A., Nobre, A. C., Bloch, G., Hyder, F., et al. (1994). Functional magnetic resonance imaging of human prefrontal cortex activation during a spatial working memory task. *Proceedings of the National Academy of Sciences of the United States of America, 91*, 8690–8694.

Moulson, M. C., Fox, N. A., Zeanah, & Nelson, C. A. (2009). Early adverse experiences and the neurobiology of facial emotional processing. *Developmental Psychology, 45*, 17–30.

Nelson, C. A. (2007). A neurobiological perspective on early human deprivation. *Child Development Perspectives, 1*, 13–18.

Nelson, C. A., Zeanah, C. H., Fox, N. A., Marshall, P. J., Smyke, A., & Guthrie, D. (2007). Cognitive recovery in socially deprived young children: The Bucharest Early Intervention Project. *Science, 318*, 1937–1940.

Owen, A. M., Downes, J. J., Sahakian, B. J., Polkey, C. E., & Robbins, T. W. (1990). Planning and spatial working memory deficits following frontal lobe lesions in man. *Neuropsychologia, 28*, 1021–1034.

Owen, A. M., Evans, A. C., & Petrides, M. (1996). Evidence for a two-stage model of spatial working memory processing within the lateral frontal cortex: A positron emission tomography study. *Cerebral Cortex, 6*, 31–38.

Owen, A. M., Iddon, J. L., Summers, B. A., & Robbins, T. W. (1997). Spatial and non-spatial working memory at different stages of Parkinson's disease. *Neuropsychologia, 35*, 519–532.

Owen, A. M., Morris, R. G., Sahakian, B. J., Polkey, C. E., & Robbins, T. W. (1996). Double dissociations of memory and executive functions in a self-ordered working memory task following frontal lobe excision, temporal lobe excisions or amygdalahippocampectomy in man. *Brain, 119*, 1597–1615.

Owen, A. M., Roberts, A. C., Polkey, C. E., Sahakian, B. J., & Robbins, T. W. (1991). Extra-dimensional versus intra-dimensional set-shifting performance following frontal lobe excisions, temporal lobe excisions, or amygdalo-hippocampectomy in man. *Neuropsychologia, 29*, 993–1006.

Owen, A. M., Sahakian, B. J., Semple, J., Polkey, C. E., & Robbins, T. W. (1995). Visuospatial short-term recognition memory and learning after temporal lobe excision, frontal lobe excision, or amygdalo-hippocampectomy in man. *Neuropsychologia, 33*, 1–24.

Parker, S. W., Nelson, C. A., & the BEIP Core Group. (2005a). An event-related potential study of the impact of institutional rearing on face recognition. *Development and Psychopathology, 17*, 621–639.

Parker, S. W., Nelson, C. A., & the BEIP Core Group. (2005b). The impact of early institutional rearing on the ability to discriminate facial expressions of emotion: An event-related potential study. *Child Development, 76*, 54–72.

Pascalis, O., Scott, L. S., Kelly, D. J., Dufour, R. W., Shannon, R. W., Nicholson, E., et al. (2005). Plasticity of Face Processing in Infancy. *Proceedings of the National Academy of Sciences of the United States of America, 102*, 5297–5300.

Passingham, R. E. (1995). *The frontal lobes and voluntary action*. Oxford, UK: Oxford University Press.

Phaf, R. H., Van der Heijden, A. H., & Hudson, P. T. (1990). SLAM: A connectionist model for attention in visual selection tasks. *Cognitive Psychology, 22*, 273–341.

Pierrot-Deseilligny, C., Rivaud, S., Gaymard, B., Muri, R., & Vermersch, A. I. (1995). Cortical control of saccades. *Annals of Neurology, 37*, 557–567.

Posner, M. I., Nissen, M. J., & Klein, R. M. (1976). Visual dominance: An information-processing account of its origins and significance. *Psychological Review, 83*, 157–171.

Robbins, T. W. (1996). Dissociating executive functions of the prefrontal cortex. *Philosophical Transactions of the Royal Society B: Biological Sciences, 351*, 1463–1471.

Robinson, C. W., & Sloutsky, V. M. (2004). Auditory dominance and its change in the course of development. *Child Development, 75*, 1387–1401.

Roid, G. H., & Miller, L. J. (1997). *Leiter International Performance Scale–Revised*. Wood Dale, IL: Stoelting.

Rutter, M. (1981). Psychological sequelae of brain damage in children. *The American Journal of Psychiatry, 138*, 1533–1544.

Rutter, M. (1998). Developmental catch-up, and deficit, following adoption after severe global early privation: English and Romanian Adoptees (ERA) Study Team. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 39*, 465–476.

Rutter, M., Dunn, J., Plomin, R., & Simonoff, E. (1997). Integrating nature and nurture: Implications of person-environment correlations and interactions for developmental psychopathology. *Development and Psychopathology, 9*, 335–364.

Saffran, J. R., Werker, J., & Werner, L. (2006). The infant's auditory world: Hearing, speech, and the beginnings of language. In R. Siegler & D. Kuhn (Eds.), *Handbook of child development* (pp. 58–108). New York: Wiley.

Saghal, A., Sahakian, B. J., Robbins, T. W., Wray, C. J., Lloyd, S., & Cook, J. H. (1991). Detection of visual memory and learning deficits in Alzheimer's disease using the Cambridge Neuropsychological Test Automated Battery. *Dementia, 2*, 150–158.

Sahakian, B. J., Morris, R. G., Evendon, J. L., Heald, A., Levy, R., & Philpot, M. (1988). A comparative study of visuospatial memory and learning in Alzheimer-type dementia and Parkinson's disease. *Brain, 111*, 695–718.

Sahakian, B. J., & Owen, A. M. (1992). Computerized assessment in neuropsychiatry using CANTAB: Discussion paper. *Journal of the Royal Society of Medicine, 85*, 399–402.

Sanchez, M. M., Hearn, E. F., Do, D., Rilling, J. K., & Herndon, J. G. (1998). Differential rearing affects corpus callosum size and cognitive function of rhesus monkeys. *Brain Research, 812*, 38–49.

Sanchez, M. M., Ladd, C. O., & Plotsky, P. M. (2001). Early adverse experience as a developmental risk factor for later psychopathology: Evidence from rodent and primate models. *Development and Psychopathology, 13*, 419–450.

Sanchez, M. M., & Pollak, S. D. (2009). Socio-emotional development following early abuse and neglect: Challenges and insights from translational research. In M. de Haan & M. R. Gunnar (Eds.), *Handbook of developmental social neuroscience* (pp. 497–520). New York: Guilford.

Shallice, T., Marzocchi, G. M., Coser, S., Del Savio, M., Meuter, R. F., & Rumiati, R. I. (2002). Executive function profile of children with attention deficit hyperactivity disorder. *Developmental Neuropsychology, 21*, 43–71.

Slotnick, S. D. (2004). Visual memory and visual perception recruit common neural substrates. *Behavioral and Cognitive Neuroscience Reviews, 3*, 207–221.

Sloutsky, V. M., & Napolitano, A. C. (2003). Is a picture worth a thousand words? Preference for auditory modality in young children. *Child Development, 74*, 822–833.

Sowell, E. R., Thompson, P. M., Holmes, C. J., Jernigan, T. L., & Toga, A. W. (1999). In vivo evidence for post-adolescent brain maturation in frontal and striatal regions. *Nature Neuroscience, 2*, 859–861.

Sweeney, J. A., Mintun, M. A., Kwee, S., Wiseman, M. B., Brown, D. L., Rosenberg, D. R., et al. (1996). Positron emission tomography study of voluntary saccadic eye movements and spatial working memory. *Journal of Neurophysiology, 75*, 454–468.

van IJzendoorn, H. W., & Juffer, F. (2006). The Emanuel Miller Memorial Lecture 2006: Adoption as intervention. Meta-analytic evidence for massive catch-up and plasticity in physical, socio-emotional, and cognitive development. *Journal of Child Psychology and Psychiatry, 27*, 1228–1245.

van IJzendoorn, M. H., Juffer, F., & Poelhuis, C. W. (2005). Adoption and cognitive development: A meta-analytic comparison of adopted and nonadopted children's IQ and school performance. *Psychological Bulletin, 131*, 301–316.

Verhulst, F. C., Althaus, M., & Versluis-denBieman, H. J. M. (1990). Problem behavior in international adoptees: III. Diagnosis of child psychiatric disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 29*, 420–428.

Verhulst, F. C., Althaus, M., & Versluis-denBieman, H. J. (1992). Damaging backgrounds: Later adjustment of international adoptees. *Journal of the American Academy of Child and Adolescent Psychiatry, 31*, 518–524.

Vincent, S. L., Khan, Y., & Benes, F. M. (1995). Cellular and co-localization of dopamine with D1 and D2 receptors in rat medial prefrontal cortex. *Synapse, 19*, 112–120.

Wechsler, D. (1998). *Wechsler Intelligence Scale for Children (WISC–III)* (3rd ed.). San Antonio, TX: Harcourt Brace.

Windsor, J., Glaze, L. E., Koga, S. F., & the BEIP Core Group. (2007). Language acquisition with limited input: Romanian institution and foster care. *Journal of Speech, Language, and Hearing Research, 50*, 1365–1381.

Zeanah, C. H., Nelson, C. A., Fox, N. A., Smyke, A. T., Marshall, P., Parker, S. W., et al. (2003). Designing research to study the effects of institutionalization on brain and behavioral development: The Bucharest Early Intervention Project. *Development and Psychopathology, 15*, 885–907.

Child Development, January/February 2010, Volume 81, Number 1, Pages 237–251

Effects of Perinatal HIV Infection and Early Institutional Rearing on Physical and Cognitive Development of Children in Ukraine

Natasha A. Dobrova-Krol, Marinus H. van IJzendoorn,
Marian J. Bakermans-Kranenburg, and Femmie Juffer
Centre for Child and Family Studies, Leiden University

To study the effects of perinatal HIV-1 infection and early institutional rearing on the physical and cognitive development of children, 64 Ukrainian uninfected and HIV-infected institutionalized and family-reared children were examined (mean age = 50.9 months). Both HIV infection and institutional care were related to delays in physical and cognitive development, with a larger effect of the rearing environment. Family care, even of compromised quality, was found to be more favorable for children's physical and cognitive development than institutional care. The impact of the quality of child care on physical and cognitive development is discussed in light of future interventions.

Early exposure to cumulative adversity may cause enduring neurobiological changes responsible for a range of developmental deficits in different domains (Nelson, 2007). The major sources of childhood adversity are poor health, economic hardship, and compromised rearing environment. The presence of perinatal HIV infection is associated with all three of them: HIV, as a medical condition, has a direct impact on physical and cognitive development of the child; as a psychosocial phenomenon, pediatric HIV infection is usually accompanied by perinatal adversities, poverty, parental substance abuse and illness, and increased risk of child neglect, maltreatment and abandonment (Klunklin & Harrigan, 2002; Steele, Nelson, & Cole, 2007).

Currently, thanks to the advances in treatment, HIV infection has turned into a chronic, subacute condition rather than an acute lethal disease (e.g., De Martino et al., 2000). At the same time, due to parental death or child abandonment, HIV infection has also become a growing reason for child institutionalization (United Nations International Children's Emergency Fund, 2007). As has been repeatedly demonstrated in previous research, institutional care jeopardizes children's optimal development (e.g., Johnson, Browne, & Hamilton-Giachritsis, 2006). Nevertheless, several studies suggest that in some cases of extremely adverse rearing circumstances well-functioning child-care institutions may offer children a better rearing environment than their own dysfunctional families (e.g., Miller et al., 2007). Therefore, the purpose of this study was to examine the separate and combined influences of HIV status and the type of the rearing environment, that is, biological families and child-care institutions, on children's stress regulation, and physical and cognitive development, as reflected by their diurnal cortisol production, linear growth, general intelligence, and social-cognitive intelligence as indicated by theory of mind.

Previous research demonstrates that HIV infection is usually accompanied by lower birth weight and physical growth delays, and may cause inhibition of cortisol production. The growth failure of HIV-infected children is usually caused by HIV replication, with the resulting immune response affecting the metabolism and consequently physical growth. Growth delays can also occur as a result of perinatal and postnatal insults, unrelated to HIV infection, such as prenatal substance exposure, prematurity, and malnutrition prevalent among children born to HIV-infected mothers, and due to the impact of other HIV clinical symptoms, concurrent diseases, and/or treatment toxicity (Bailey,

---

This research was supported by research grants to the second and third authors from the Netherlands Organization for Scientific Research (NWO SPINOZA Prize; NWO VIDI Grant 452-04-306). F.J. is supported by Wereldkinderen.

Correspondence concerning this article should be addressed to Marinus H. van IJzendoorn, Centre for Child and Family Studies, Leiden University, PO Box 9555, NL-2300RB Leiden, Netherlands. Electronic mail may be sent to vanijzen@fsw.leidenuniv.nl.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8101-0016

Kamenga, Nsuami, Nieburg, & St Louis, 1999; European Collaborative Study, 2003). Furthermore, HIV may cause inhibition of the hypothalamic–pituitary–adrenal system functioning due to the direct effect of the virus on adrenal gland, or as a result of concurrent diseases and/or medical treatment (e.g., Marik, Kiminyo, & Zaloga, 2002).

Cognitive deficits in HIV-infected children can be attributed to structural central nervous system abnormalities and progressive encephalopathy caused by HIV, health condition, and stage of the disease progress (see Wachsler-Felder & Golden, 2002, for a review). Developmental delays in HIV-infected children can, however, also be caused by adverse rearing environment (Brown, Lourie, & Pao, 2000). In fact, several studies suggest that adverse rearing environment may have equal or even larger effects on the development of children than the infection itself (Coscia, Christensen, Henry, Wallston, & Radcliffe, 2001; Smith et al., 2006).

An average family provides children with what is referred to as the average expectable environment. Depending on the child's age, it encompasses a range of species-specific elements, such as protective, stable caregiving, and socialization, and opens opportunities for exploration of the world (Bowlby, 1998; Cicchetti & Valentino, 2006; Hartmann, 1958). The provision of the average expectable environment is an important prerequisite for children's normal development. For children reared in HIV-impacted families, poverty, parental substance abuse, and serious illness are compromising factors falling outside the range of the average rearing environment and leading to adverse developmental outcomes (e.g., Blanchette, Smith, King, Fernandes-Penney, & Read, 2002; De Bellis, 2005; Fishkin et al., 2000; Gottlieb & Blair, 2004).

As to institutional rearing, it clearly falls outside the range of the average expectable environment due to structural neglect that is embedded in the organization and functioning of child-care institutions (van IJzendoorn, 2008): Its regimented nature, high child-to-caregiver ratio, multiple shifts, and frequent change of caregivers almost inevitably deprive children of continuous and reciprocal interactions with stable caregivers, necessary to respond to their developmental needs. Compromised family rearing environment as well as institutional care can lead to physical growth delays (Skuse, Reilly, & Wolke, 1994; van IJzendoorn, Bakermans-Kranenburg, & Juffer, 2007), stress dysregulation (Gunnar & Vazquez, 2006), lower IQ (Nobel, McCandliss, & Farah, 2007; van IJzendoorn, Luijk, & Juffer, 2008), and delayed false belief understanding (Cicchetti, Rogosch, Maughan, Toth, & Bruce, 2003; Pears & Fisher, 2005; Tarullo, Bruce, & Gunnar, 2007; Yagmurlu, Berument, & Celimli, 2005).

The link between compromised rearing environment and developmental deficits may operate through child–caregiver interactions. Studies suggest that even in the condition of structural institutional neglect the quality of caregiving plays a leading role in the children's development. Thus, Smyke et al. (2007) found that for institution-reared children, the caregiving quality was related to three of six developmental outcomes, including cognitive development. Based on adoptive parents' retrospective reports on the standards of institutional care, Castle et al. (1999) established that individualized care in institutions had the strongest positive effect on adoptees' IQ. Intervention studies also support the importance of child–caregiver relationships, demonstrating that even a modest increase of visual, tactile, and auditory interaction between a child and a caregiver (Hakimi-Manesh, Mojdchi, & Tashakkori, 1984), or some extra untutored human care (Hunt, Mohandessi, Ghodssi, & Akiyama, 1976) results in improved physical and mental development of institution-reared children (for a review and meta-analysis of interventions with institutionalized children, see Bakermans-Kranenburg, van IJzendoorn, & Juffer, 2008).

Physical surroundings represent another aspect of the rearing environment influencing child development (directly and via caregivers; see Evans, 2006, for a review). Morison, Ames, and Chisholm (1995) showed that developmental delays in children with early institutional experience were less severe in the presence of toys and worse when children were dirty or soiled while in institutional care. Studies of family-reared children consistently demonstrated that a stimulating and well-organized physical environment allowing free exploration of developmentally adequate objects, toys and books, and language stimulation by caregivers have a positive effect on children's development (e.g., Bradley, 1985). In the current study, we explored how these factors were related to the development of HIV-infected children.

According to Gunnar's (2001) classification, the institutions involved in our study were characterized by the second level of institutional privation, providing adequate nutrition and health care, however, lacking stimulation and stability in child–caregiver relationships (Dobrova-Krol, van IJzendoorn, Bakermans-Kranenburg, Cyr, & Juffer, 2008). This being the case, the question arose whether family care even of compromised quality, as in the case of HIV-impacted families, facilitates the development of HIV-infected children better than institutional care. Or will perhaps institutions that provide fairly clean environments, good medical care, and nutrition, such as those where the present study was conducted, be good enough to facilitate the development of children better than the troubled family environment that HIV-impacted parents provide? And to what extent does the presence of HIV undermine the development of the institution-reared children as compared to their noninfected institution-reared peers? To address these questions, we compared stress regulation, and physical and cognitive development of HIV-infected children with about the same level of immune control over the infection, as reflected by their CD4 T-lymphocyte count (McMichael & Rowland-Jones, 2001), in different rearing environments: families versus institutions. To examine the effect of HIV infection on physical and cognitive development with respect to the type of the rearing environment, we also included family- and institution-reared children without the HIV condition.

We predicted the following: (a) both HIV infection and institutional rearing would lead to stress dysregulation and delayed physical and cognitive development, with an expected larger effect of the rearing environment than of HIV infection, and (b) even compromised family care was expected to be more favorable for the child's development than institutional care, but quality of care within the institution was predicted to be associated with level of cognitive development.

Method

Participants

Participants were 64 children: 13 HIV-infected institution-reared children (mean age = 52.28 months, *SD* = 12.99, range = 34.78–74.16), 16 uninfected institution-reared children (mean age = 48.14 months, *SD* = 9.72, range = 35.11–66.73), 16 HIV-infected family-reared children, (mean age = 52.01 months, *SD* = 14.78, range = 30.18–71.56), and 19 uninfected family-reared children (mean age = 51.44 months, *SD* = 9.77, range = 37.48–67.06). The groups did not differ with respect to age and gender. Data on the background characteristics of children are presented in Table 1. All HIV-infected children were born to seropositive mothers and acquired the infection perinatally. Diagnosis of HIV-1 infection was based on positive viral culture of polymerase chain reaction assay performed after the neonatal period on at least two occasions over 2 months. In the group of uninfected institution-reared children, 5 had perinatal hypoxic neurological conditions (PHNC), whereas there were no cases of PHNC in the three remaining groups. Because previous studies have demonstrated that PHNC are related to altered physical and cognitive development and stress regulation (Ellis et al., 2001; Van Handel, Swaab, De Vries, & Jongmans, 2007), we decided to exclude these children from further analysis. Moreover, one of the HIV-infected family-reared children was excluded because of clear signs of fetal alcohol syndrome not reported in the medical records, resulting in an effective sample of 58 children. The power of the current

Table 1
Descriptive Statistics for Family-Reared Versus Institution-Reared Children

	Family-reared children				Institution-reared children			
	HIV−		HIV+		HIV−		HIV+	
	n^a	*M* (*SD*)	n^a	*M* (*SD*)	n^a	*M* (*SD*)	n^a	*M* (*SD*)
Age of mother (years)	17	32.12 (5.93)	14	32.47 (5.68)	11	30.18 (8.73)	8	30.38 (5.21)
Family income	12	234.75_a (112.29)	13	149.54_b (43.10)				
Age of child (months)	19	51.44 (9.77)	16	52.01 (14.78)	16	48.14 (9.72)	13	52.28 (12.99)
Weight-for-age at birth	16	-0.32_a (0.99)	15	-1.15_{ab} (1.21)	15	-0.81_{ab} (0.61)	13	-1.36_b (0.56)
CD4 T-lymphocyte count			14	913.21 (459.91)			13	1,147.69 (375.01)
		n (%)		*n* (%)		*n* (%)		*n* (%)
Perinatal complications	19		16		16	5 (37.5)	13	
Antiretroviral therapy			16	6_a (37.5)			13	12_b (92.3)

Note. Mean values in the same row that do not share subscripts differ at $p < .05$.
[a] Number of cases with available data.

design with two factors (HIV and rearing environment; $N = 58$ children) and an expected large effect size ($f = .40$) was .82; for the multivariate analyses including one covariate (birth weight; see the following) the power was .85.

Institution-reared children. Institution-reared children were recruited from four Children's Homes in Ukraine. The following selection criteria were applied: (a) age between 3 and 6 years old, (b) admission to institutional care within the 1st year of life, (c) no genetic syndromes (e.g., Down syndrome), (d) no evidence of fetal alcohol syndrome in the medical records, and (e) permanent residence in residential care institutions since admission. These selection criteria substantially limited the number of children who could participate in the study. On average, the mean age at admission to institutional care of the uninfected children was 1 month ($SD = 1$, range = 0–3), and they had been living in institutions for 47 months ($SD = 9$, range = 35–65). The mean age at admission to institutional care of the HIV-infected children was also 1 month ($SD = 2$, range = 0–7). One child who was admitted at 7 months of age previously had lived in a hospital and was cared for by the hospital staff. HIV-infected children had been living in institutional care for 51 months on average ($SD = 14$, range = 35–74).

The data on the history of institutionalization showed that in the group of uninfected institution-reared children, 9 were admitted to institutional care because of poverty, 2 because of family disruption, 4 because one or both parents were in prison, and 1 child was an orphan. Since admission to institutional care, 7 of the uninfected children were transferred from one institution to another. For 1 child, information about the number of transfers between institutions was missing. Seven uninfected children remained in occasional contacts with their parents or family members. Among the uninfected institution-reared children, 2 had siblings living in the same institution; the siblings did not participate in our study.

In the group of HIV-infected institution-reared children, 1 child was admitted to institutional care due to poverty, 1 child due to family disruption, and 5 children because one or both parents were in prison; 6 of the HIV-infected institution-reared children were orphans. Five HIV-infected children were once transferred from one child-care institution to another. Three of the HIV-infected children remained in occasional contacts with their biological family members. HIV-infected children had no siblings residing in the same institution.

Family-reared children. Family-reared children were recruited from local Ukrainian kindergartens, schools, and clinics where routine health checks take place. Caregivers of HIV-infected family-reared children were contacted through a nongovernmental organization providing psychosocial support to HIV-infected people in the region. Some of the contacted caregivers refused to participate in the study mainly because several measures (see the following) involved the video-recording of activities at home; some parents refused to participate because of the seemingly complicated procedure of diurnal salivary cortisol sampling, which along with the strict selection criteria limited our sample size. Children were selected according to the following criteria: (a) age between 3 and 6 years old, (b) living in biological families, (c) no genetic syndromes (e.g., Down syndrome), (d) no evidence of fetal alcohol syndrome in the medical records, and (e) no previous history of institutionalization.

All uninfected children were reared in two-parent biological families with at least one employed parent. In two cases, the excessive use of alcohol by fathers was reported. There were no reported cases of criminal records among the parents of uninfected family-reared children. Among HIV-infected family-reared children, 8 were living in two-parent biological families, 4 were reared by single mothers, and 3 by their single biological grandmothers as their parents were unable to take care of the children due to drug abuse. In eight families, at least one parent was using alcohol and/or drugs and in one family both parents were abusing alcohol. In three families, both parents were unemployed. In four families, at least one parent had a criminal record and in one case both parents had criminal records. In two cases, data on criminal records of the fathers were missing. Both uninfected and HIV-infected family-reared children were raised in families with low to middle income, but the monthly income of the HIV-impacted families was significantly lower than of families without HIV, $t(23) = 2.54$, $p < .05$. We tested whether income of the families as a proxy for family socioeconomic status (SES) impacted the outcome measures of the family-reared children, but no significant difference emerged ($.15 < p < .46$).

Procedure

For all children enrolled in the study, informed consent was obtained; for the children in the children's homes from the local department of the Ministry of Health, and for the children in the family-reared group from their parents or primary caregivers. All children were invited for a laboratory assessment procedure. Institution-reared children were accompanied by their "favorite" caregiver, as determined through preliminary informal interviews with children and caregivers. If a favorite caregiver was difficult to identify, the person who spent most of the time with a child and knew him or her best was invited. Family-reared children were accompanied by their primary caregiver. During the laboratory assessment procedure, the children underwent a physical examination (height, weight, and head circumference) and were administered an abbreviated, nonverbal cognitive performance test (Snijders-Oomen Nonverbal Intelligence Test–Revised [SON–R]) and a Theory of Mind (ToM) task. Because of limited time and resources, assessment of the rearing environment was possible only for the HIV-infected children and was conducted during home visits in the case of family-reared children and during visits to childcare institutions for institution-reared children.

Measures

Child HIV-1 health status. Following the assessment procedure, the children's medical records from the regional AIDS center were reviewed in order to obtain the following data: immunological status according to the Centers for Disease Control and Prevention (CDC) classification, cluster of differentiation 4 (CD4) T-lymphocyte counts that reflect the level of immune control over the infection (McMichael & Rowland-Jones, 2001), and duration and type of antiretroviral treatment. CD4 cell counts most proximal to the time of assessment were used (range = 1–3 months). Preliminary analysis revealed no significant difference between the HIV-infected family-reared and institution-reared children on CD4 cell counts, $t(27) = -1.48$, $p = .15$ (see Table 1). Out of 20 children subjected to the treatment, 18 received the same antiretroviral medications. Table 1 shows that 37.5% of family-reared children and 92.3% of institution-reared children were receiving antiretroviral therapy at the time of the assessment; no significant difference between family-reared and institution-reared children was found on duration of antiretroviral treatment $t(15) = -0.57$, $p = .58$. We tested whether the children receiving antiretroviral therapy differed from the nontreated HIV-infected children on all outcome measures, but no significant difference emerged ($.15 < p < .94$).

Physical growth. Data on physical growth were collected from the children's medical records. Measurements of current height, weight, and head circumference of all children were conducted by a trained research assistant during the laboratory visit. For the purposes of the present study, we used birth weight as a marker for prenatal adversities (e.g., Nordstrom-Klee, Delaney-Black, Covington, Ager, & Sokol, 2002; Shankaran et al., 2004). Current height, weight, and head circumference were highly correlated ($rs > .50$), and because faltering of height serves as a good indicator of long term chronic adversities (e.g., Miller, 2005), current height-for-age scores were used as index for physical growth. Height-for-age (HAZ) was calculated with the software program Epi Info™, Version 3.3.2 using the sex-specific 2000 CDC reference database (Dean et al., 2002).

Diurnal salivary cortisol sampling. A six-sample protocol was followed. Saliva samples were collected on a typical day from the institution-reared children by an institutional nurse and from the family-reared children by their parent. The standard saliva collection protocol and assay procedure for determining cortisol concentration in the saliva samples was applied (see for details, Bakermans-Kranenburg, van IJzendoorn, Mesman, Alink, & Juffer, 2008; Dobrova-Krol et al., 2008). Due to the low concentration of saliva within the cotton swabs, 5 HIV-infected and 5 uninfected family-reared children had missing data at one or two of the six sampling points. In these cases, missing cortisol values were generated with log curve estimation analyses using individual sampling times as the independent variable, resulting in 52 children with complete cortisol data. Three HIV-infected children and 1 uninfected family-reared child who had missing data at more than two sampling points as well as 1 HIV-infected family-reared child who refused to cooperate were not included in the analysis involving diurnal cortisol. Another uninfected family-reared child was excluded from the analyses involving diurnal cortisol because he was ill on the day of saliva sampling. Diurnal cortisol scores were log 10-transformed prior to analyses (Azar et al., 2004).

In order to assess the overall secretion of cortisol from awakening until bedtime the "area under the curve with respect to ground" (AUCg) was computed (Pruessner, Kirschbaum, Meinlschmid, & Hellhammer, 2003). Since the AUCg was related to the total time that the children were awake (from awakening till bed time), we corrected the AUCg for children's total time of being awake. In the group of HIV-infected institution-reared children, two outliers on the AUCg were assigned a raw score on the offending variable that differed 1 unit from the next most extreme score in the distribution (Tabachnik & Fidell, 2007).

Cognitive performance. The SON-R was used to assess the children's cognitive performance (Tellegen, Winkel, Wijnberg-Williams, & Laros, 1998). The SON–R is a nonverbal intelligence test for children between 2.5 and 7 years of age that does not require the use of spoken or written language. The test consists of six nonverbal subtests that focus on two different types of abilities: visual-spatial abilities (meaningful pictures) and abstract reasoning (nonmeaningful pictures). Previous research showed that the nonverbal SON–R test is well suited for use with children of ethnic minorities in the Netherlands (Tellegen & Laros, 1993) and in other countries (e.g., Australia: Jenkinson, Roberts, Dennehy, & Tellegen, 1996; China: Zhang, Gong, Sun, & Tian, 1997). In our study, we selected two of the most reliable subtests of the SON–R to assess the children's cognitive performance (Tellegen & Laros, 1993): a visual-spatial performance Patterns subtest and an abstract reasoning Analogies subtest; average internal consistency of the subtests was .76. Convergent validity of the SON–R IQ with the Wechsler Intelligence Scale for Children–Revised Full Scale Intelligence Quotient and Performance Intelligence Quotient was .80, and with the Raven's matrices .74 (Tellegen & Laros, 1993). In our sample, the internal consistency of the used subtests as estimated by Cronbach's alpha was .81. Total cognitive performance score was calculated with the SON–R computer program on the basis of the two subtests.

Theory of Mind. To assess the children's ability to comprehend false belief, we used one of the False Belief tasks: the classic Unexpected-Transfer Task (Baron-Cohen, Leslie, & Frith, 1985). The Unexpected-Transfer Task has a number of advantages when testing children without or with very little productive language skills due to its easily understood storyline supplemented by the use of props, simple vocabulary, and its nonverbal response mode (Baron-Cohen, 1992; Garfield, Peterson, & Perry, 2001). The Unexpected-Transfer Task demonstrated good test–retest reliability, with a kappa of .62 (Hughes et al., 2000). The task was acted out with two dolls, Sally and Anne, each with a different colored basket and a ball, and involved unseen displacement of the object, in which Sally puts a ball in her basket and then leaves the scene. While Sally is away and cannot watch, Anne takes the ball out of Sally's basket and puts it into her box. Sally then returns and the children are asked where they think she will look for her ball. To pass the task children were to answer that Sally will look inside her basket before realizing that her ball is not there, and thus predict an action based on an attributed false belief. The names of the dolls were changed into the culturally appropriate ones.

Rearing environment. The Home Observation for Measurement of the Environment (HOME) inventory was employed to assess the rearing environments of HIV-infected family- and institution-reared children. The Early Childhood HOME inventory is designed to measure the quality and quantity of stimulation and support available to a child in the rearing environment (Bradley et al., 1993). The information needed to score the HOME was obtained during a 45- to 90-min visit to the place where the child lived, during a time when the child and the child's primary caregiver were present and awake. The semistructured interview and observation were conducted with minimal obtrusiveness and allowed observed participants to act normally (Caldwell & Bradley, 2003). The Child Care Early Childhood version of the HOME (CC–EC–HOME) designed for use between 3 and 6 years of age was employed in the study. It contains 58 items clustered into eight subscales: Learning Materials, Language Stimulation, Physical Environment, Caregiver Responsivity, Academic Stimulation, Modeling, Variety of Experience, and Acceptance. The inventory was translated into Russian and the interviewers were trained by the first author. In order to calculate interrater reliabilities, a second observer coded 93% of the CC–EC–HOME observations: Intraclass correlation coefficients for all the subscales scores and the total HOME scores ranged from .73 to .94.

Background characteristics. Univariate analyses of variance (ANOVAs) and chi-square tests showed no significant differences between the four groups on age of the biological mother, child gender, or child age. Differences on the variables related to child background characteristics are presented in Table 1.

Results

Preliminary Analysis

Preliminary analyses did not reveal an association between gender and any of the outcome variables, that is, height at the assessment, diurnal cortisol production, and cognitive performance. Birth weight was significantly related to current height ($p < .01$), marginally related to cognitive performance on SON–R ($p = .06$), and not related to ToM ($p = .11$). We used birth weight as a covariate in our analyses involving height and SON–R. Current age of the child was significantly related only to ToM ($p < .01$) and was taken into account in analyses involving this variable. Mean values and standard deviations of height, diurnal cortisol production, cognitive performance on SON–R, and percentages of children who passed the ToM task are presented in Table 2.

Differences Between Uninfected and HIV-Infected Family- and Institution-Reared Children on Height, Stress Regulation, and Cognitive Performance

Height. Figure 1, based on the height-for-age mean values with standard errors from birth to 48 months of age, shows the development of height of the four groups across infancy and early childhood. Although both groups of HIV-infected children had lower supine length at birth than their uninfected counterparts, this difference was significant only between uninfected family-reared children and HIV-infected institution-reared children ($p < .01$). Figure 1 shows a normal pattern of linear growth of the uninfected family-reared children remaining within 1 *SD* (range = 0.29 to 0.87) of the reference population. For the other three groups, a faltering of linear growth after birth was observed at different periods of their development: the height of the HIV-infected family-reared children remained within 1 *SD* (range = −0.97 to 0.08) of the reference population, the height of the uninfected institution-reared children ranged from −2.09 to −0.75, and the height of the HIV-infected institu-

Table 2

Mean Values and Standard Deviations for Height, Diurnal Cortisol Production, and Cognitive Performance on SON–R, as a Function of Rearing Environment and HIV Status, Controlling for Birth Weight and Percent of Children Who Passed the ToM Task

	Family-reared children				Institution-reared children					
	HIV−		HIV+		HIV−		HIV+			
	n	*M* (*SD*)	*n*	*M* (*SD*)	*n*	*M* (*SD*)	*n*	*M* (*SD*)	*F*	χ^2
Height	16	0.12_a (0.78)	15	-0.61_{ab} (1.06)	11	-0.89_{bc} (0.72)	13	-1.86_c (1.30)	7.02**	
Cortisol[a]	15	0.45_a (0.18)	12	0.47_{ab} (0.27)	11	0.63_b (0.15)	13	0.40_a (0.17)	2.80*	
SON–R	16	98.94_a (19.59)	15	79.07_b (16.90)	11	69.73_{bc} (21.28)	13	64.00_c (14.32)	8.59**	
ToM[b]	18	61.1_a	16	25.0_b	11	18.2_b	13	23.1_b		8.33*

Note. Mean values in the same row that do not share subscripts differ at $p < .05$. Height = height-for-age Z score at the assessment day; SON–R = Snijders-Oomen Nonverbal Intelligence Test–Revised; ToM = Theory of Mind.

[a] Diurnal production of cortisol computed with AUCg (area under the curve with respect to ground) formula. [b] Unexpected-Transfer Task.

*$p < .05$. **$p < .01$.

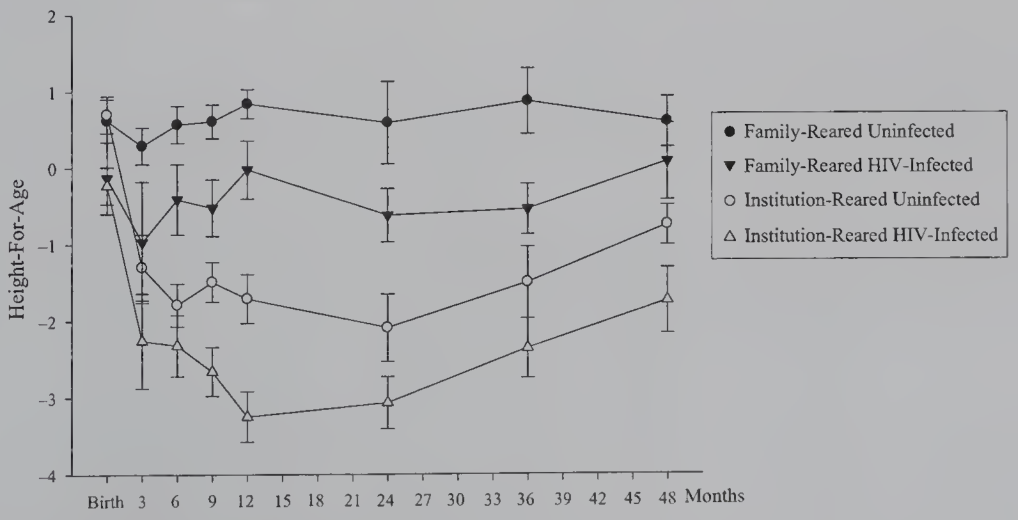


Figure 1. Height-for-age (Z scores, mean, standard error) of uninfected and HIV-infected family- and institution-reared children.
Note. At 36 months, *N* = 47 (family-reared uninfected = 8, family-reared infected = 16, institution-reared uninfected = 10, institution-reared infected = 13); at 48 months, *N* = 35 (family-reared uninfected = 8, family-reared infected = 11, institution-reared uninfected = 7, institution-reared infected = 9).

tion-reared children ranged from −3.25 to −1.74. At the time of the assessment, all four groups had mean height-for-age scores within 2 *SD* from the reference population (see also Table 2).

Stress regulation: Diurnal cortisol production. Figure 2 illustrates diurnal cortisol curves of uninfected and HIV-infected family- and institution-reared children. Table 2 shows the mean values of the total diurnal cortisol production for the four groups of children.

Cognitive performance: SON–R and ToM. As Table 2 shows, cognitive performance on SON–R of the uninfected family-reared children was in the average range (M = 98.94, SD = 19.59), performance of HIV-infected family-reared children was in the borderline deficiency range (M = 79.07, SD = 16.90), and performance of uninfected institution-reared children (M = 69.73, SD = 21.28) and HIV-infected children (M = 64.00, SD = 8.59) fell in the range of mental deficiency according to the SON–R norms (Tellegen et al., 1998). Table 2 presents the percentages of children who passed the ToM task.

Multivariate Analyses

In Table 3, bivariate correlations between the control and outcome variables in the whole sample and as a function of the rearing environment and HIV infection are presented. To assess the difference between the groups as a function of the rearing environment and HIV status, we conducted a multivariate analysis of covariance (MANCOVA) on current height, diurnal cortisol production, and cognitive performance, controlling for birth weight. Significant overall effects were found for rearing environment, $F(3, 44) = 10.23$, $p < .01$, partial $\eta^2 = .41$, and for HIV status, $F(3, 44) = 3.01$, $p = .04$, partial $\eta^2 = .17$. Subsequent ANCOVAs showed main effects for rearing environment and HIV status for both height and cognitive performance and a

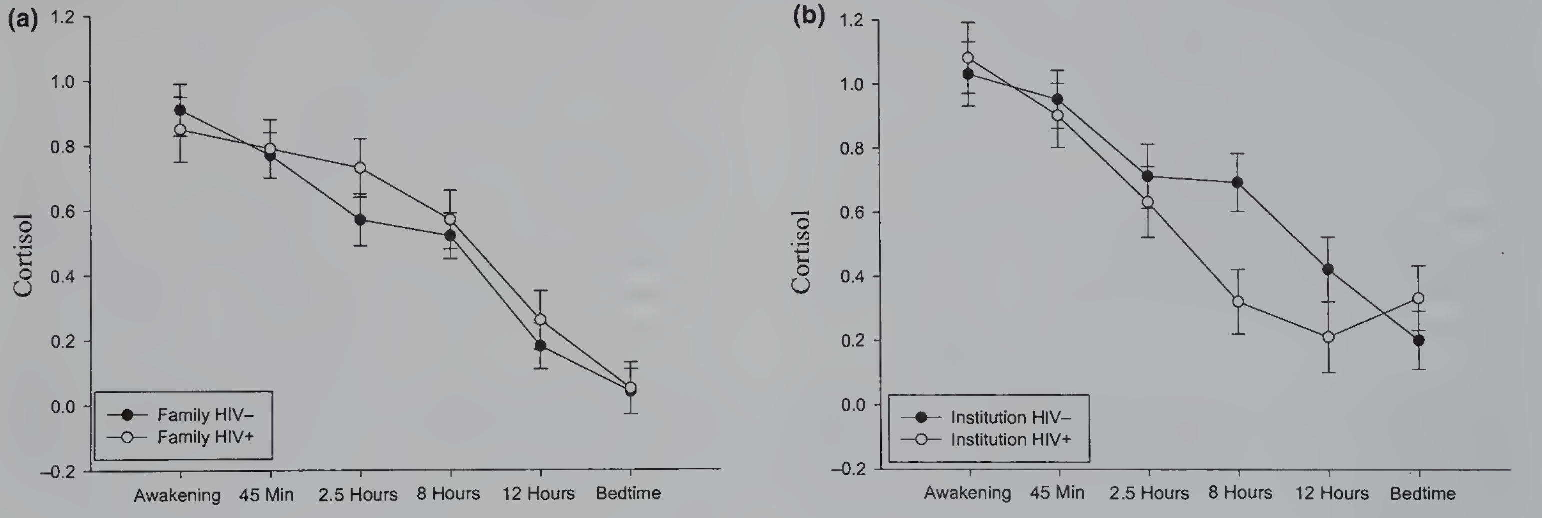


Figure 2. Diurnal cortisol values (nmol/L, log-transformed) of uninfected and HIV-infected family-reared children (a) and institution-reared children (b).

Table 3
Intercorrelations for the Outcome Variables Across and Within Uninfected Versus HIV-Infected, Family- Versus Institution-Reared Groups

												Uninfected group						Family-reared group
	Total						HIV-infected group						Institution-reared group					
	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
1.WAZ	—						—	−.16	.23	−.04	.25	.25	—	−.03	.28	.04	.29	.23
2. Age	.03	—					.06	—	−.25	.03	−.11	.31	.09	—	−.19	.05	.01	.35*
3. HAZ	.38**	.02	—				.29	−.17	—	−.47*	.27	.49**	.60**	.15	—	−.03	.22	.40*
4. Cortisol	.12	−.15	−.08	—			.08	.06	.01	—	−.28	−.36	.38	−.37	.01	—	−.03	.16
5. SON–R	.25	.12	.41**	.00	—		−.01	.12	.39*	.12	—	.26	.01	.17	.20	.27	—	.08
6 ToM	.22	.36**	.44**	−.09	.24	—	.43*	.45	.34	.06	.03	—	.09	.32	.37	−.41*	.24	—

Note. WAZ = weight-for-age Z score at birth; HAZ = height-for-age Z score at the assessment day; SON–R = Snijders-Oomen Nonverbal Intelligence Test–Revised; ToM = Theory of Mind.
*$p < .05$. **$p < .01$.

significant interaction effect for diurnal cortisol production (see Table 2). A priori contrasts revealed that uninfected family-reared children had significantly better scores on current height than both institution-reared groups, and performed significantly better on the SON–R test than all other groups. No significant difference on current height and SON–R between the HIV-infected family-reared children and uninfected institution-reared children was found. HIV-infected institution-reared children had significantly lower current height and SON–R scores than both family-reared groups, but they did not differ significantly from the uninfected institution-reared group (see Table 2). Diurnal production of cortisol was significantly higher in the group of uninfected institution-reared children as compared to uninfected family-reared children and HIV-infected institution-reared children, but uninfected institution-reared children did not differ significantly from HIV-infected family-reared children.

For ToM, chi-square analysis based on the extension of the Fisher's exact test for larger contingency tables and for small sample sizes (Verbeek & Kroonenberg, 1985) demonstrated significant differences between the groups, $\chi^2(3, N = 58) = 8.33$, $p = .04$, with family-reared uninfected children outperforming the other three groups of children. To control for age, we used a median split of the sample on age and compared the children's ToM performance among younger (< 47.86 months) and older (> 47.86 months) children. We found that in the group of younger children the success rate was 43% for uninfected family-reared children and 17% for HIV-infected institution-reared children; none of the HIV-infected family-reared and uninfected institution-reared children passed the task. Among older children the success rate was 73% for uninfected family-reared children, 67% for uninfected institution-reared children, 50% for HIV-infected family-reared children, and 29% for HIV-infected institution-reared children. Separate chi-square analysis conducted to compare the performance of uninfected and HIV-infected family- and institution-reared children in the younger and older groups on ToM demonstrated a significant effect for the younger children, $\chi^2(3, N = 29) = 7.58$, Fisher exact $p = .04$, again with family-reared uninfected children outperforming the other three groups of children, and no effect for the older children, $\chi^2(3, N = 29) = 3.62$, Fisher exact $p = .35$.

Cognitive Development and Environmental Characteristics

To compare the rearing environments of HIV-infected family- and institution-reared children, one-way ANOVAs on HOME subscales and HOME total scores were conducted (see Table 4; HOME scores were not available for the non-HIV-infected participants; see the Method section). To obtain an accurate picture of the magnitude of the differences between the two groups on the various HOME scales, independent of sample size, we also present effect sizes, as reflected by Cohen's *d*. According to Cohen (1988), $d = 0.20$ can be interpreted as a small effect size, $d = 0.50$ as medium, and $d = 0.80$ as large. The rearing environment of the HIV-infected institution-reared children was scored significantly

Table 4
Mean Values, Standard Deviations, and Correlations With Cognitive Performance on SON–R Scores for HOME Scales of HIV-Infected Children as a Function of Rearing Environment

	Total (*N* = 29)		Family-reared children (*N* = 16)		Institution-reared children (*N* = 13)			
Scale	*M* (*SD*)	*r*	*M* (*SD*)	*r*	*M* (*SD*)	*r*	*F*	*d*
Learning Materials	6.41 (2.57)	.29	6.13 (2.96)	.31	6.77 (2.05)	.54	0.44	−0.25
Language Stimulation	5.69 (1.39)	.54**	6.00 (1.67)	.50**	5.31 (0.86)	.52	1.83	0.52
Physical Environment	4.79 (1.61)	−.18	3.81 (1.60)	.20	6.00 (0.00)	.25	24.11**	−1.94
Responsivity	5.93 (1.71)	.12	6.00 (1.75)	.02	5.85 (1.73)	.24	0.06	0.07
Academic Stimulation	4.10 (1.18)	.13	4.19 (1.38)	.08	4.00 (0.91)	.19	0.18	0.16
Modeling	5.14 (1.27)	.23	5.44 (1.46)	.10	4.77 (0.93)	.25	2.05	0.55
Variety	6.41 (1.84)	−.02	5.63 (2.09)	.24	7.38 (0.77)	.23	8.23**	−1.11
Acceptance	3.17 (0.89)	.42*	3.63 (0.89)	.15	2.62 (0.51)	.48	13.31**	2.84
Total	41.66 (8.02)	.28	40.81 (10.21)	.30	42.69 (4.21)	.66**	0.39	−0.24

Note. SON–R = Snijders-Oomen Nonverbal Intelligence Test–Revised; HOME = Home Observation for Measurement of the Environment.
**p* < .05. ***p* < .01.

higher than the environment of HIV-infected family-reared children on the Physical Environment subscale, $F(1, 27) = 24.11$, $p < .01$ ($d = -1.94$), and the Variety subscale, $F(1, 27) = 8.23$, $p < .01$ ($d = -1.11$), and significantly lower on the Acceptance subscale, $F(1, 27) = 13.31$, $p < .01$ ($d = 2.84$). No difference was found on the HOME total score, $F(1, 27) = 0.39$, $p = .54$ ($d = -0.24$).

To examine how various aspects of the rearing environment as reflected by the HOME subscales were associated with the cognitive performance of HIV-infected children, we computed the correlations between the HOME subscales and SON–R scores (see Table 4). For the whole group of HIV-infected children higher SON–R scores were related to more Language Stimulation ($r = .54$), and more Acceptance ($r = .42$). For HIV-infected family-reared children higher SON–R scores were related to more Language Stimulation ($r = .50$). For HIV-infected institution-reared children, higher SON–R scores were related to higher total HOME scores ($r = .66$). We did not find any significant relations between the HOME subscales and current height, diurnal cortisol production, and ToM performance.

Discussion

Both HIV infection and institutional care are related to delayed physical growth and poor cognitive performance. The impact of the rearing environment on physical growth and cognitive performance on SON–R was, however, greater than the impact of HIV infection. Family care, even of compromised quality, was found to be more favorable for children's physical growth and cognitive development than institutional care. HIV-infected children reared in disadvantaged families showed better physical and cognitive development not only in comparison to HIV-infected children, but also compared to noninfected relatively healthy children reared in institutions providing good quality physical environments. In families as well as in institutions, better quality of the rearing environment was associated with higher levels of cognitive performance of HIV-infected children. We found elevated diurnal cortisol production in uninfected institution-reared children, and lower levels of diurnal cortisol production in their HIV-infected counterparts. Finally, uninfected children reared in families performed significantly better on the ToM task than all other groups.

The design of the current study provides a unique opportunity to examine how perinatal HIV infection and early rearing environment affect stress regulation, physical growth, and cognitive development of children in an ethnically homogeneous group. Moreover, to control for major differences in the duration and type of institutional care, we assessed children who were admitted to institutional care at 1 month after birth on average and who permanently resided in child-care institutions with about the same level of institutional privation.

Physical Growth

The association between HIV status and lower birth weight and physical growth delays has been well documented in previous studies (e.g., Bailey et al., 1999; European Collaborative Study, 2003), and was replicated in our sample. Archival data showed that delays in physical development of HIV-infected children were evident already at birth and persisted through the course of the child's development as compared to noninfected children raised in similar rearing environments. However, physical growth delays of HIV-infected children reared in families were less substantial in comparison to not only HIV-infected but also uninfected institution-reared children, who received adequate nutritional provision and health care, and whose physical environment was not significantly different from the environment of their family-reared counterparts as revealed by the HOME total score.

Stress Regulation

Our finding of an interaction effect between HIV status and type of care on total diurnal cortisol production indicates that the institutional environment, as a source of repeated daily intermittent stress, may cause an elevation of cortisol production for uninfected children, as also documented in other studies addressing stress regulation of children with institutional experience (e.g., Gunnar, Morison, Chisholm, & Schuder, 2001). In the case of HIV-infected children, it may contribute to a weakened immune system functioning that allows HIV to have a greater inhibiting impact on cortisol production as compared to HIV-infected family-reared children. However, since family- and institution-reared children had similar CD4 T-lymphocyte counts (reflecting the level of immune control over the infection), an alternative and perhaps more plausible explanation is related to antiretroviral treatment that the majority of the HIV-infected institution-reared children and less than half of the family-reared children were receiving, which may also inhibit production of cortisol (e.g., Marik et al., 2002).

Cognitive Performance on SON–R

The cognitive performance of the uninfected family-reared children was in the average range, whereas in the HIV-infected family-reared children it was in the borderline deficiency range, and in both uninfected and HIV-infected institution-reared children it fell in the range of mental deficiency (Tellegen et al., 1998). For institutionalized children, the impact of the infection on cognitive functioning appeared to be not very substantial. In fact, not all aspects of cognitive functioning may be equally affected by the presence of HIV. For example, along with normal performance of HIV-infected children on various measures of IQ, lower levels of performance have been found for executive functioning and processing speed (Koekkoek, De Sonneville, Wolfs, Licht, & Geelen, 2007), visual-spatial and time orientation tests (Tardieu et al., 1995), and motor development (Blanchette et al., 2002). The SON–R test used in our study was not meant to detect impairments in executive functioning.

At the same time, a substantial discrepancy in cognitive performance between family-reared children with and without HIV was observed: It was more than 3 times larger than between institution-reared children with and without HIV. This discrepancy may be attributed to the difference in the rearing circumstances in the HIV-infected and uninfected families. Whereas uninfected children were from relatively normally functioning low-to-middle class families, HIV-infected children were growing up in more problematic family environments. The relatively small difference in cognitive performance of institution-reared children may be explained by the fact that both uninfected and HIV-infected children were raised in mixed groups in the same institutions, and their rearing circumstances were nearly identical. Alternatively, it may reflect a general floor effect in non-neurologically impaired groups.

On average, the institution-reared children were more than 22 IQ points behind the family-reared children, and the gap between HIV-infected institution-reared children and their family-reared counterparts was more than 15 IQ points. Examination of the rearing environments of HIV-infected children living in families and institutions with the HOME inventory revealed no significant differences on the HOME total score between those groups. Interestingly, scores on the individual HOME scales indicate that in comparison to the rearing environment of HIV-impacted families, child-care institutions were providing more toys, books, and games that facilitate learning; the physical environment of the institutions was safer, offered more space, and was perceptually more appealing. Finally, the care arrangements in the institutions provided more variety for the children. However, despite these benefits, HIV-infected institution-reared children showed greater cognitive delay than their family-reared counterparts. Further examination of the individual scales of the HOME inventory revealed that caregivers of the HIV-infected family-reared children were more accepting. Basically, institutional care was offering a better physical environment, whereas families secured more supportive child–caregiver interaction. Our findings suggest that the presence of a primary caregiver, and family care, even of compromised quality, promotes cognitive development of HIV-infected children above institutional care. Importantly, we found that correlations between cognitive performance and different aspects of the rearing environment reflected by the HOME scales were of considerable strength—even if they were not always statistically significant due to small subsample sizes. Also, the overall quality of care within institutions was associated with the cognitive performance of HIV-infected children; however, fairly adequate physical aspects of the rearing environment appeared to be insufficient to facilitate normative cognitive development.

Theory of Mind

More than half of the uninfected family-reared children passed the ToM task assessing the acquisition of false belief understanding, which was more than two times more than in the remaining three groups of children. This difference was especially evident among the younger children. This finding confirms that the average expectable environment facilitates earlier acquisition of ToM. However, although HIV-impacted families secured better child–caregiver interaction than in institutions, it was not sufficiently better to facilitate early ToM development, which requires from caregivers a certain level of sensitivity, fine tuning and mind-mindedness as well as the ability to focus on the child's thoughts and feelings (Fonagy, Gergely, & Target, 2007). In the same vein, another study addressing the development of ToM in family- and institution-reared children demonstrated significantly better results for family-reared children from middle-SES families and failed to demonstrate statistically significant differences between children reared in low SES families and in institutions (Yagmurlu et al., 2005).

Limitations and Future Directions

The major limitation of the current study is its modest sample size that calls for replication of our findings in other countries and institutional settings. However, the power of the study was adequate (> .80) and sufficient to find significant differences of the expected magnitude. The design of the study did not allow for random assignment of children to conditions. Because of the scarce information on perinatal development, we were only able to account for possible effects of birth weight as a proxy for perinatal insults. Besides, children in the four settings may have been systematically different in their genetic makeup, and further studies are needed to extend our understanding of how such genetic differences might contribute to the development of institution-reared children, making them less or more vulnerable to the negative effects of structural neglect (Rutter, 2006). Finally, the group of uninfected family-reared children was different from the other three groups in that these families most likely provided the average expectable environment. On the one hand, conclusions regarding the effects of HIV status on developmental outcomes should remain preliminary until direct comparison with a group of high-risk families without HIV-infected children is made. On the other hand, it does allow assessing differences in the effects of the average expectable environment and institutional care among uninfected children. It should also be noted that income differences between the two types of families were not associated with children's growth parameters, cortisol production, and cognitive development.

Practical Implications

Our study has important practical implications, especially concerning the development and implementation of intervention efforts in child-care institutions, which is the important strength and impact of this study. Our results indicate that it seems to be the quality of child–caregiver relationships and not so much the presence of HIV infection or the quality of the physical environment that contributed to the (delays in) growth and cognitive development of the children in our study. This finding points to the importance of comprehensive but focused intervention efforts in child-care institutions. Renovation of premises and equipping them with toys and learning materials, which has become a popular form of intervention in Eastern Europe, is certainly valuable, but our findings suggest that they may not be sufficient to decrease the gap between the institutional environment and the environment necessary for normative child development.

Of course, family-type of care is the best alternative for institution-reared children. Recent findings of the Bucharest Early Intervention Project (BEIP) in which children were randomly assigned to foster families or institutional care demonstrated that by the age of 4, children in foster care were scoring almost 10 points higher on IQ tests than the children left in the orphanages, and those children who left the orphanages before 2 years of age saw an almost 15-point increase (Nelson et al., 2007). However, much time may pass before in Ukraine, with about 112,000 children under the age of 17 years reared in institutions (State Institute for Family and Youth Development, 2007), emerging alternative care and especially foster care for children with special needs and HIV will substitute institutional care. Unfortunately, HIV-related stigma and fears are still dominating among the general population. Some of the caregivers who participated in our study admitted that they had to conceal their involvement with HIV-infected children from their families. Not surprisingly, children with HIV remain the least preferred candidates for adoption or foster care. Therefore, intervention programs ameliorating the impact of institutionalization are of high importance for those children who are lacking the alternatives of growing up in family care. We believe that in such cases institutions offering good nutritional support and health care, especially important when adherence to the treatment regime is required, may serve at least as a temporary resort for the children. However, changing the structural neglect by enhancing the stability and the quality of the child–caregiver relationship should be a precondition and the major target of the intervention efforts (St. Petersburg-USA Orphanage Research Team, 2008).

Thus, our study fulfills an important pilot function, increasing our knowledge about the physical and cognitive development of an emerging special-needs group of HIV-infected children reared in child-care institutions. Especially now, when the number of HIV-infected children in Ukraine and elsewhere in the world is growing, we hope that the examination of possible risk and protective factors affecting their development will contribute to a better understanding of their developmental processes, and to practical recommendations that ultimately lead to an improvement of their quality of life.

References

Azar, R., Zoccolillo, M., Paquette, D., Quiros, E., Baltzer, F., & Tremblay, R. E. (2004). Cortisol levels and conduct disorder in adolescent mothers. *Journal of the American Academy of Child and Adolescent Psychiatry, 43*, 461–468.

Bailey, R. C., Kamenga, M. C., Nsuami, M. J., Nieburg, P., & St. Louis, M. E. (1999). Growth of children according to maternal and child HIV, immunological and disease characteristics: A prospective cohort study in Kinshasa, Democratic Republic of Congo. *International Journal of Epidemiology, 28*, 532–540.

Bakermans-Kranenburg, M. J., van IJzendoorn, M. H., & Juffer, F. (2008). Earlier is better: A meta-analysis of 70 years of intervention improving cognitive development in institutionalized children. *Monographs of the Society for Research of Child Development, 73*, 279–293.

Bakermans-Kranenburg, M. J., van IJzendoorn, M. H., Mesman, J., Alink, L. R. A., & Juffer, F. (2008). Effects of an attachment-based intervention on daily cortisol moderated by DRD4: A randomized control trial on 1-3-year-olds screened for externalizing behavior. *Development and Psychopathology, 20*, 805–820.

Baron-Cohen, S. (1992). The theory of mind hypothesis of autism: History and prospects of the idea. *The Psychologist, 5*, 9–12.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a 'theory of mind? *Cognition, 21*, 37–46.

Blanchette, M., Smith, M. L., King, S., Fernandes-Penney, A., & Read, S. (2002). Cognitive development in school age children with vertically transmitted HIV infection. *Developmental Neuropsychology, 21*, 223–241.

Bowlby, J. (1998). *Attachment and loss: Vol. 2: Separation.* London: Pimlico.

Bradley, R. H. (1985). Social-cognitive development and toys. *Topics in Early Childhood Special Education, 5*, 11–30.

Bradley, R. H., Whiteside, L., Caldwell, B. M., Casey, P. H., Kelleher, K., Pope, S., et al. (1993). Maternal IQ, the home environment, and child IQ in low birthweight premature children. *International Journal of Behavioral Development, 16*, 61–74.

Brown, L. K., Lourie, K. J., & Pao, M. (2000). Children and adolescents living with HIV and AIDS: A review. *Journal of Child Psychology and Psychiatry, 41*, 81–96.

Caldwell, B. M., & Bradley, R. H. (2003). *Home observation for measurement of the environment.* Little Rock: University of Arkansas at Little Rock.

Castle, J., Groothues, C., Bredenkamp, D., Becket, C., O'Connor, T., & Rutter, M. et al. (1999). Effects of qualities of early institutional care on cognitive attainment. *American Journal of Orthopsychiatry, 69*, 424–437.

Cicchetti, D., Rogosch, F. A., Maughan, A., Toth, S. L., & Bruce, J. (2003). False belief understanding in maltreated children. *Development and Psychopathology, 15*, 1067–1091.

Cicchetti, D., & Valentino, K. (2006). An ecological transactional perspective on child maltreatment: Failure of the average expectable environment and its influence upon child development. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology risk, disorder, and adaptation* (Vol. 3, 2nd ed., pp. 129–201). New York: Wiley.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Earlbaum.

Coscia, J. M., Christensen, B. K., Henry, R. R., Wallston, K., & Radcliffe, J. (2001). Effects of home environment, socioeconomic status, and health status on cognitive functioning in children with HIV-1 infection. *Journal of Pediatric Psychology, 26*, 321–329.

Dean, A. G., Arner, T. G., Sunki, G. G., Friedman, R., Lantinga, M., & Sangam, S. et al. (2002). *Epi Info, a database and statistics program for public health professionals.* Atlanta, GA: Centers for Disease Control and Prevention.

De Bellis, M.D. (2005). The psychobiology of neglect. *Child Maltreatment, 10*, 150–172.

De Martino, M., Tovo, P., Balducci, M., Galli, L., Gabiano, C., Rezza, G., et al. (2000). Reduction in mortality with availability of antiretroviral therapy for children with perinatal HIV-1 infection. *JAMA, 284*, 190–197.

Dobrova-Krol, N., van IJzendoorn, M., Bakermans-Kranenburg, M., Cyr, C., & Juffer, F. (2008). Physical growth delays and stress dysregulation in stunted and non-stunted Ukrainian institution-reared children. *Infant Behavior and Development, 31*, 539–553.

Ellis, M., Shrestha, L., Shrestha, P. S., Manandhar, D. S., Bolam, A. J., & de L Costello, A.M. (2001). Clinical predictors of outcome following mild and moderate neonatal encephalopathy in term newborns in Kathmandu, Nepal. *Acta Paediatrica, 90*, 316–322.

European Collaborative Study. (2003). Height, weight and growth in children born to mothers with HIV-1 infection in Europe. *Pediatrics, 111*, e52–e60.

Evans, G. W. (2006). Child development and the physical environment. *Annual Review of Psychology, 57*, 255–284.

Fishkin, P. E., Armstrong, F. D., Routh, D. K., Harris, L., Thompson, W., Miloslavich, K., et al. (2000). Brief report: Relationship between HIV infection and WPPSI-R performance in preschool age children. *Journal of Pediatric Psychology, 25*, 347–351.

Fonagy, P., Gergely, G., & Target, M. (2007). The parent-infant dyad and the construction of the subjective self. *Journal of Child Psychology and Psychiatry, 48*, 288–328.

Garfield, J. L., Peterson, C. C., & Perry, T. (2001). Social cognition, language acquisition and the development of theory of mind. *Mind and Language, 16*, 494–541.

Gottlieb, G., & Blair, C. (2004). How early experience matters in intellectual development in the case of poverty. *Prevention Science, 5*, 245–252.

Gunnar, M. R. (2001). Effects of early deprivation. In C. A. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (pp. 617–629). Cambridge, MA: MIT Press.

Gunnar, M. R., Morison, S. J., Chisholm, K., & Schuder, M. (2001). Salivary cortisol levels in children adopted from Romanian orphanages. *Development and Psychopathology, 13*, 611–628.

Gunnar, M. R., & Vazquez, M. (2006). Stress neurobiology and developmental psychopathology. In D. Cicchetti & D. Cohen (Eds.), *Developmental psychopathology, developmental neuroscience* (Vol. 2, 2nd ed., pp. 533–577). New York: Wiley.

Hakimi-Manesh, Y., Mojdchi, H., & Tashakkori, A. (1984). Short communication: Effects of environmental enrichment on the mental and psychomotor development of orphanage children. *Journal of Child Psychology and Psychiatry, 25*, 643–650.

Hartmann, H. (1958). *Ego psychology and the problem of adaptation*. New York: International Universities Press.

Hughes, C., Adlam, A., Happe, F. G. E., Jackson, J., Taylor, A., & Capsi, A. (2000). Good test-retest reliability for standard and advanced false-belief tasks across a wide range of abilities. *Journal of Child Psychology and Psychiatry, 41*, 483–490.

Hunt, J. M., Mohandessi, K., Ghodssi, M., & Akiyama, M. (1976). The psychological development of orphanage reared infants: Interventions with outcomes. *Genetic Psychology Monographs, 94*, 177–226.

Jenkinson, J., Roberts, S., Dennehy, S., & Tellegen, P. (1996). Validation of the Snijders-Oomen Nonverbal Intelligence Test–Revised 2.5-7 for Australian children with disabilities. *Journal of Psychoeducational Assessment, 14*, 276–286.

Johnson, R., Browne, K., & Hamilton-Giachritsis, C. (2006). Young children in institutional care at risk of harm: A review. *Trauma, Violence and Abuse, 7*, 34–60.

Klunklin, P, & Harrigan, R. C. (2002). Child-rearing practices of primary caregivers of HIV-infected children: An integrative review of the literature. *Journal of Pediatric Nursing, 17*, 289–296.

Koekkoek, S., De Sonneville, L. M., Wolfs, T. F., Licht, R., & Geelen, S. P. (2007). Neurocognitive function profile in HIV-infected school-age children. *European Journal of Paediatric Neurology, 12*, 290–297.

Marik, P. E., Kiminyo, K., & Zaloga, G.P. (2002). Adrenal insufficiency in critically ill HIV infected patients. *Critical Care Medicine, 30*, 1267–1273.

McMichael, A. J., & Rowland-Jones, S. L. (2001). Cellular immune responses to HIV. *Nature, 410*, 980–987.

Miller, L. C. (2005). *The handbook of international adoption medicine. A guide for physicians, parents, and providers*. Oxford, UK: Oxford University Press.

Miller, L. C., Chan, W., Litvinova, A., Rubin, A., Tirella, L., & Cermak, S. (2007). Medical diagnoses and growth of children residing in Russian orphanages. *Acta Paediatrica, 96*, 1765–1769.

Morison, S. J., Ames, E. W., & Chisholm, K. (1995). The development of children adopted from Romanian orphanages. *Merrill-Palmer Quarterly, 41*, 411–430.

Nelson, C. (2007). A neurobiological perspective on early human deprivation. *Child Development Perspectives, 1*, 13–18.

Nelson, C., Zeanah, C. H., Fox, N. A., Marshall, P. J., Smyke, A. T., & Guthrie, D. (2007). Cognitive recovery in socially deprived young children: The Bucharest Early Intervention Project. *Science, 318*, 1937–1940.

Nobel, K. G., McCandliss, B. D., & Farah, M. J. (2007). Socioeconomic gradients predict individual differences on neurocognitive abilities. *Developmental Science, 10*, 464–480.

Nordstrom-Klee, B., Delaney-Black, V., Covington, C., Ager, J., & Sokol, A. (2002). Growth from birth onwards of children prenatally exposed to drugs: A literature review. *Neurotoxicology and Teratology, 24*, 481–488.

Pears, K. C., & Fisher, P. (2005). Emotion understanding and theory of mind among maltreated children in foster care: Evidence of deficits. *Development and Psychopathology, 17*, 47–65.

Pruessner, J. C., Kirschbaum, C., Meinlschmid, G., & Hellhammer, D.H. (2003). Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. *Psychoneuroendocrinology, 28*, 916–931.

Rutter, M. (2006). *Genes and behavior. Nature-nurture interplay explained*. Oxford, UK: Blackwell.

Shankaran, S., Das, A., Bauer, C. R., Bada, H. S., Lester, B., Wright, L., et al. (2004). Association between patterns of maternal substance use and infant birth weight, length, and head circumference. *Pediatrics, 114*, e226–e234.

Skuse, D., Reilly, S., & Wolke, D. (1994). Psychosocial adversity and growth during infancy. *European Journal of Clinical Nutrition, 48*, 113–130.

Smith, R., Malee, K., Leighty, R., Brouwers, P., Mellins, C., Hittelman, C., et al. (2006). Effects of perinatal HIV infection and associated risk factors on cognitive development among young children. *Pediatrics, 117*, 851–862.

Smyke, A. T., Koga, S.F., Johnson, D.E., Fox, N.A., Marshall, P.J., Nelson, C. A., et al. (2007). The caregiving context in institution reared and family reared infants and toddlers in Romania. *Journal of Child Psychology and Psychiatry, 48*, 210–218.

State Institute for Family and Youth Development (2007). *National Report on Implementation of the Decisions of the Final Outcome Document of the United Nations General Assembly Special Session on Children (2002) and Action Plan "A World Fit for Children."* Kiev: Ministry of Family, Youth, and Sport of Ukraine.

Steele, R. G., Nelson, T. D., & Cole, B. P. (2007). The psychosocial adjustment of children and adolescents with HIV: Review from a socio-ecological perspective.

Journal of Developmental and Behavioral Pediatrics, 28, 58–69.

St. Petersburg-USA Orphanage Research Team. (2008). The effects of early social-emotional and relationship experience on the development of young orphanage children. *Monographs of the Society of Research in Child Development, 73* (Serial No. 291).

Tabachnik, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). New York: Allyn & Bacon.

Tardieu, M., Mayaux, M., Seibel, N., Funck-Bretano, I., Straub, E., Tegles, J., et al. (1995). Cognitive assessment of school-age children infected with maternally transmitted human immunodeficiency virus Type 1. *Journal of Pediatrics, 126*, 375–379.

Tarullo, A. R., Bruce, J., & Gunnar, M. R. (2007). False belief and emotion understanding in post-institutionalized children. *Social Development, 16*, 57–78.

Tellegen, P., & Laros, J. (1993). The construction and validation of a nonverbal test of intelligence: The revision of the Snijders-Oomen tests. *European Journal of Psychological Assessment, 9*, 147–157.

Tellegen, P. J., Winkel, M., Wijnberg-Williams, B. J., & Laros, J. A. (1998). *Snijders-Oomen Nonverbal Intelligence Test, SON-R 2.5-7, Manual and research report*. Lisse, Netherlands: Swets & Zeitlinger.

United Nations International Children's Emergency Fund. (2007). Children and AIDS: A stocktaking report. Retrieved February 15, 2008, from http://data.unaids.org/pub/Report/2007/20060116_stocktaking_report.pdf

Van Handel, M., Swaab, H., De Vries, L. S., & Jongmans, M. J. (2007). Long-term cognitive and behavioral consequences of neonatal encephalopathy following perinatal asphyxia: A review. *European Journal of Pediatrics, 166*, 645–654.

van IJzendoorn, M. (2008). *Opvoeding over de grens. Gehechtheid, trauma en veerkracht*. [Parenting across borders. Attachment, trauma and resilience]. Amsterdam: Boom Academic.

van IJzendoorn, M. H., Bakermans-Kranenburg, M. J., & Juffer, F. (2007). Plasticity of growth in height, weight and head circumference: Meta-analytic evidence of massive catch-up of children's physical growth after adoption. *Journal of Developmental and Behavioral Pediatrics, 28*, 334–343.

van IJzendoorn, M. H., Luijk, M., & Juffer, F. (2008). IQ of children growing up in children's homes: A meta-analysis on IQ delays in orphanages. *Merrill-Palmer Quarterly, 54*, 341–366.

Verbeek, A., & Kroonenberg, P. (1985). A survey of algorithms for exact distributions of test statistics in r x c contingency tables with fixed margins. *Computational Statistics and Data Analysis, 3*, 159–185.

Wachsler-Felder, J. L., & Golden, C. J. (2002). Neuropsychological consequences of HIV in children: A review of current literature. *Clinical Psychology Review, 22*, 441–462.

Yagmurlu, B., Berument, S. K., & Celimli, S. (2005). The role of institution and home contexts in theory of mind development. *Journal of Applied Developmental Psychology, 26*, 521–537.

Zhang, H., Gong, W., Sun, Y., & Tian, X. (1997). Snijders-Oomen Nonverbal Intelligence Test–Revision (SON–R) in China. *Psychological Science China, 20*, 97–103.

Child Development, January/February 2010, Volume 81, Number 1, Pages 252–269

The Differential Impacts of Early Physical and Sexual Abuse and Internalizing Problems on Daytime Cortisol Rhythm in School-Aged Children

Dante Cicchetti
University of Minnesota

Fred A. Rogosch
University of Rochester

Megan R. Gunnar
University of Minnesota

Sheree L. Toth
University of Rochester

The impact of early physical and sexual abuse (EPA/SA) occurring in the first 5 years of life was investigated in relation to depressive and internalizing symptomatology and diurnal cortisol regulation. In a summer camp context, school-aged maltreated (n = 265) and nonmaltreated (n = 288) children provided morning and late afternoon saliva samples on 5 consecutive days. Child self-report and adult observer reports of child internalizing and depressive symptoms were obtained. Children experiencing EPA/SA and high depressive or internalizing symptoms uniquely exhibited an attenuated diurnal decrease in cortisol, indicative of neuroendocrine dysregulation. These results were specific to EPA/SA rather than later onset physical or sexual abuse or early occurring neglect or emotional maltreatment.

What happens to biological and psychological functioning when there are severe dysfunctions in the early caregiving environment? The study of child maltreatment provides researchers with an excellent opportunity to address this question guided by preclinical research on the impact of early life stress on the development of fear and neuroendocrine dysregulation and by human research on maltreatment, depression, and activity of the hypothalamic–pituitary–adrenocortical (HPA) system. The present report focuses on the impact of physical and sexual abuse (often in combination with neglect) during the first 5 years of life as an example of severe dysfunction of the early caregiving environment.

Among the array of difficulties exhibited by maltreated children, one of the most prevalent and widely documented is an increased risk for internalizing problems and depression in childhood and major depressive disorder in adulthood (Gibb, Wheeler, Alloy, & Abramson, 2001; Kaplow & Widom, 2007; Keiley, Howe, Dodge, Bates, & Pettit, 2001; Kim & Cicchetti, 2006; Manly, Kim, Rogosch, & Cicchetti, 2001; Toth, Manly, & Cicchetti, 1992; Widom, DuMont, & Czaja, 2007). A variety of factors have been posited to mediate the impact of child maltreatment on depression. These include subtype of maltreatment, genetic vulnerability, the quality of children's relationships with their mothers, social support, stressful life events, attributional styles, social competence, self-esteem, and stress hormone activity (Cicchetti & Valentino, 2006). Regarding stress hormones, it has been argued that for those at genetic risk of affective disorder, elevations in stress hormones during episodes of abuse may shape the development of more reactive fear and neuroendocrine systems, and hence increase the risk of developing internalizing disorders (Heim & Nemeroff, 2001).

The animal literature provides support for this hypothesis (Heim, Owen, Plotsky, & Nemeroff, 1997). Chronic elevations in glucocorticoids (cortisol in humans and other primates), produced by the HPA axis, increase the production of fear-mediating neurochemicals in the central nucleus of the amygdala (Makino, Gold, & Schulkin, 1994). In nonhuman primates, individuals' carrying gene

---

This research was supported by funding from the National Institute on Drug Abuse (DA12903, DA17741) and the Spunk Fund Inc.

Correspondence concerning this article should be addressed to Dante Cicchetti, Institute of Child Development, University of Minnesota, 51 East River Road, Minneapolis, MN 55455. Electronic mail may be sent to cicchett@umn.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8101-0017

variants associated with risk for depression in abused children (e.g., Caspi et al., 2003) exhibit larger cortisol responses to psychosocial stressors (Barr et al., 2004; Sanchez et al., 2005). Furthermore, a large body of animal studies shows that adverse parental care during early development increases fearful, anxious behavior and shapes increased reactivity in neurobiological systems involved in defensive behaviors and physiological stress reactions (see Gunnar & Quevedo, 2007; Gunnar & Vazquez, 2006, for reviews). In rodents and nonhuman primates, this has been demonstrated through studies of fear behavior and activity of the HPA axis (Levine, 2005; Maestripieri, 1999; Meaney & Szyf, 2005; Sanchez, McCormack, & Maestripieri, in press).

The neurobiological signature of adverse early care has been shown to be highly consistent at the systems and molecular levels to neurobiological findings in the study of adult mood disorder (see Heim, Newport, Mletzko, Miller, & Nemeroff, 2008; Heim, Plotsky, & Nemeroff, 2004). It has been argued that these early effects reflect both the dependence of the very young on maternal care for survival and the rapid development of the brain during the early period of life (Gunnar, 2003; Gunnar & Vazquez, 2006; Heim & Nemeroff, 2001). Thus, these animal studies suggest that inadequate or abusive parenting early in life persistently sensitizes the neurocircuits that are involved in the regulation of stress and emotion, while also suggesting that individuals at genetic risk for depression may be the most susceptible to these early experience effects (Heim & Nemeroff, 2001).

The HPA system has figured prominently in research linking early adverse care to later internalizing and mood disorders for a number of reasons. First, this neuroendocrine system is frequently dysregulated in adults with clinical depression (Holsboer, 2000). Second, both the releasing hormone (corticotropin-releasing hormone [CRH]) and the final hormonal product of this system (cortisol in humans) play critical roles in brain development, affecting neurotrophic factors and regulating gene transcription (Gunnar & Vazquez, 2006). Extremely adverse early care has been shown in animal models to increase cell death in the infant brain (Zhang et al., 2002). During adverse early care, CRH produced in the hippocampus disturbs its normal development (Fenoglio, Brunson, & Baram, 2006), and chronic early stress produces increases in CRH production in the amygdala and a shift in the expression of receptors toward those supporting heightened fear and anxiety (Sanchez, Ladd, & Plotsky, 2001). Finally, heightened fear and activation of the HPA system as a function of early adverse care are then expected to produce an increase in allostatic load over the life course, increasing vulnerability to mood disorders in response to stressors experienced later in development (McEwen, 2003; McEwen & Stellar, 1993).

Models of the impact of early caregiving adversity typically focused on *increased* reactivity of the HPA axis as a likely outcome. However, particularly as the research has been extended to human and nonhuman primates, it has become clear that a more accurate prediction is that the HPA axis will likely exhibit evidence of *dysregulation* and that this dysregulation may involve both hyper- and hypofunctioning (Heim et al., 2004, Heim et al., 2008; McEwen, 2000). Although not yet fully understood, a number of factors have been shown to affect whether elevated or suppressed activity is noted. The HPA axis tends to down-regulate in response to prolonged hyperactivation; thus, in response to chronic stressors first hyper- and later hypoactivity of the axis may be noted (see for review, Miller, Chen, & Zhou, 2007). With regard to basal cortisol, chronic stress and negative emotionality tend to be associated with slight elevations in the late afternoon and evening, but often lower than typical levels in the early morning (e.g., DeSantis et al., 2007; Gunnar & Quevedo, 2007). Thus, whether elevations or suppressions in baseline levels are reported may depend on time of measurement both in terms of time of day and time since the onset of a chronic stressor. Furthermore, in part because of counterregulation within the axis, cortisol measures (baseline or reactivity) may appear normal, even when assessments higher up in the axis (e.g., adrenocorticotropic hormone [ACTH] response to CRH challenge or pharmacological probes of feedback sensitivity) may reveal dysfunction (see, e.g., Heim et al., 2008). Finally, there is some speculation that individuals at genetic risk for antisocial personality disorder will exhibit attenuation or down-regulation of cortisol activity (Susman, 2006), while individuals at risk for depression will exhibit elevations or hyperactivity of the HPA axis following adverse early life conditions (Heim & Nemeroff, 2001; Heim et al., 2004). Overall, predicting that the measures of the HPA system will reflect dysregulation of the axis, rather than predicting that any given measure will be elevated or suppressed, reflects the current status of knowledge in the field. This is especially true in research on young children from vulnerable populations, where researchers are often limited to assessing basal measures of cortisol at only one or a few times during the day.

With these caveats in mind, we review the literature on adverse early caregiving, including maltreatment, and its association with HPA axis dysregulation in humans. Just as not all individuals who experience abuse and neglect as young children grow up to develop clinical depression (Cicchetti & Rogosch, 1996), not all go on to exhibit patterns of HPA axis dysregulation. There is increasing evidence that adults who were abused as children exhibit dysregulation of the HPA axis only or primarily if they have major depression and that adults with major depression in the absence of early child abuse do not show evidence of neuroendocrine dysregulation (Heim et al., 2004; Heim et al., 2008). The results of Heim, Newport, Bonsall, Miller, and Nemeroff (2001) are consistent with studies of children and adolescents with depression. Nonmaltreated children with early onset depression rarely display evidence of regulatory dysfunction of the HPA axis (Feder et al., 2004). Evidence of dysregulation has, however, been noted for depressed children who also were sexually abused and suicidal (Dahl & Ryan, 1996) and for children who were abused and who have chronic posttraumatic stress disorder (Carrion et al., 2002; DeBellis et al., 1999). Similarly, two previous studies of maltreated children have shown that evidence of diurnal dysregulation of the HPA axis (slightly lower morning and slightly elevated afternoon cortisol levels) tends to be observed only in maltreated children who are also suffering from depression (Hart, Gunnar, & Cicchetti, 1996; Kaufman, 1991).

Evidence of dysregulation extends to studies of pharmacological challenge. Specifically, Kaufman et al. (1997) administered a CRH challenge test to a sample of depressed abused, depressed nonabused, and normal control group children. Compared to the depressed nonabused and the normal control group children, the depressed abused children exhibited significantly greater peak, total, and net ACTH secretion post-CRH infusion. Finally, Cicchetti and Rogosch (2001) found that compared to symptomatic nonmaltreated children, maltreated children with clinical-level internalizing problems were distinguished by higher morning, afternoon, and average daily cortisol levels across a 5-day period, suggestive of HPA axis dysregulation.

While dysregulation of the HPA axis appears to be associated with maltreatment for children who are high in internalizing problems or who are clinically depressed, it remains unclear whether this is due to the type, severity, or developmental timing of the maltreatment.

To date, the relevant published studies have either focused on physical or sexual abuse (DeBellis et al., 1999; Heim & Nemeroff, 2001; Kaufman et al., 1997) or have not examined a specific subtype of maltreatment. However, most maltreated children experience multiple types of maltreatment, and the previous studies have not attempted to differentiate among them (Cicchetti & Barnett, 1991; Cicchetti & Rizley, 1981). In particular, abuse and neglect frequently co-occur. Moreover, these maltreatment subtypes have different onsets, occur with varying severities, and can recur (Bolger, Patterson, & Kupersmidt, 1998; English et al., 2005; Manly et al., 2001). In addition, children experience maltreatment during different developmental periods. Thus, maltreatment is a heterogeneous phenomenon, and researchers and clinicians must consider the diversity in the type(s), perpetrator(s), onset, frequency, chronicity, and timing of maltreatment experiences. Even more striking, few such investigations also include a measure of timing, a critical variable in neuroendocrinology (Tarullo & Gunnar, 2006). In particular, inclusion of the temporal dimension is especially important because animal studies have revealed that timing is a crucial factor. Indeed the empirical evidence gleaned from rodent and nonhuman primate investigation indicates that adverse or maltreating (i.e., neglectful and abusive) care early in an animal's life has long-lasting effects. Attempts to translate the animal research on the criticality of timing have utilized a range of time periods (i.e., infancy through adolescence) to specify what constitutes early adverse care. Thus, while there is some evidence that HPA axis dysregulation is associated with maltreatment in the context of depression, little is known about whether the type, timing, or duration of maltreatment is relevant to these findings.

Because, based on the animal literature, dysfunction in the caregiving system early in life is believed to increase the risk of depression in genetically vulnerable individuals (e.g., Heim et al., 2004; Heim et al., 2008), the literature on maltreatment and risk of depression might provide guidance. However, similar problems exist in reviewing the literature on maltreatment and internalizing disorders. Most studies examining maltreatment and depression risk have not examined maltreatment subtypes. Those studies that have examined subtypes often focus on one (e.g., sexual abuse) and not others. Not surprisingly, the results of these studies are inconsistent; some investigations reveal that physically abused and sexually abused children have the greatest likelihood of developing depression, whereas others have shown this is the more likely outcome for neglected and emotionally maltreated children (Cicchetti & Valentino, 2006). Likewise, few prospective studies have examined the interaction between timing and type(s) of maltreatment on depression outcome.

One study that did address all subtypes of maltreatment found that serious neglect experienced during the preschool period was associated with a higher rate of internalizing symptoms (Manly et al., 2001). Another investigation indicated that the experience of harsh physical punishment before age 5 was related to higher teacher ratings of internalizing behaviors in kindergarten children (Keiley et al., 2001). Unfortunately, all types of maltreatment were not addressed in this study. In a study of adults, however, Kaplow, Dodge, Amaya-Jackson, and Saxe (2005) discovered that an earlier age of sexual abuse onset predicted higher levels of anxiety years later. Similarly, Kaplow and Widom (2007) found that individuals whose age of onset of maltreatment (i.e., physical abuse or neglect) was prior to 5 years of age reported higher levels of depressive and anxiety symptoms later in life than did individuals with a later onset of maltreatment (i.e., sexual abuse). These studies suggest that early-occurring maltreatment may be particularly important in the experience of later depression and anxiety. However, limitations in the range of subtypes examined within the studies or the extent to which specific subtypes were adequately represented in the sample preclude definitive conclusions.

In the current investigation, we operationalized early maltreatment as any incident of physical abuse, sexual abuse, neglect, or emotional maltreatment occurring within the first 5 years of life. We chose the infancy, toddler, and preschool periods as our index of early maltreatment for two primary reasons. First, national epidemiological studies of child maltreatment highlight this period as a critical time for the emergence of neglect and abuse (U.S. Department of Health and Human Services, Administration on Children, Youth, and Families, 2000). Second, despite the fact that the neural systems underlying stress reactivity/regulation and internalizing problems undergo a prolonged period of development, components of the limbic system such as the hippocampus, amygdala, and the pathways to the prefrontal cortex from these structures develop rapidly over the first years of life (Thompson & Nelson, 2001).

We also chose to focus on physical and sexual abuse experienced early in the child's life (in the first 5 years). In order to extend the work of Heim (see, e.g., Heim et al., 2008), we utilized prospective reporting of abuse, an earlier age of abuse onset, and evidence for depressive and internalizing symptomatology and cortisol dysregulation occurring earlier in development during the school-age years.

Hypotheses and Research Questions

This study was designed to examine the hypothesis that children who were physically abused, sexually abused, or both prior to age 5 (early physical abuse or sexual abuse [EPA/SA]) and who expressed high depressive and internalizing symptoms would exhibit a more dysregulated pattern of diurnal cortisol activity (less decline, flatter slope) than would other maltreated children (no early physical or sexual abuse [NEPA/SA]) or nonmaltreated comparison (NC) children. Prior to testing this hypothesis, we first examined whether EPA/SA children were at higher risk for developing high depressive or internalizing symptoms than were NEPA/SA or NC children. We also explored the specificity of the finding by contrasting cortisol patterns of the high depressive or internalizing symptom EPA/SA children and subsets of the other maltreated children. First, we examined the contrast with high- and low-symptom children who experienced other forms of maltreatment (neglect, emotional maltreatment) *prior* to age 5. Next, we examined the contrast with high- and low-symptom children whose physical or sexual abuse began after age 5. Finally, we examined whether other parameters of maltreatment (onset, recency, chronicity, diversity, and perpetrators) might be alternative processes that could account for the atypical diurnal regulation for high-symptom EPA/SA children.

Method

Participants

Participants included 553 children aged 7–13 ($M_{age} = 10.02$, $SD = 1.87$) who attended a summer day camp research program designed for school-aged low-income children. The sample included both maltreated children ($n = 265$) and nonmaltreated children ($n = 288$). All of the children were in their 1st year of camp attendance. The children were not in foster care placements and were residing with their biological mothers. Demographic characteristics are presented in Table 1 for the EPA/SA, NEPA/SA, and NC groups. The proportion of boys differed between the NC and both the EPA/SA and NEPA/SA groups, $\chi^2(2) = 17.86$, $p < .001$; thus, gender was controlled in analyses. The groups did not differ in age, race/ethnicity, or in family history of receiving welfare benefits.

Recruitment and Classification Procedures

Parents of all children provided informed consent for their child's participation, as well as for examination of Department of Human Services (DHS) records pertaining to the family. Children in the maltreated group had been identified by the county DHS as having experienced child abuse or neglect, and the sample was representative of the children in families receiving services from the DHS. A recruitment liaison from DHS contacted eligible maltreating families, explained the study, and if parents were interested, released their names to the project team for recruitment. Families were free to choose whether or not to participate. Comprehensive searches of DHS records were completed and maltreatment information was coded utilizing operational criteria from the maltreatment nosology specified in the Maltreatment Classification System (MCS; Barnett, Manly, & Cicchetti, 1993), as discussed below.

Consistent with national demographic characteristics of maltreating families (Third National Incidence Study of Child Abuse and Neglect [NIS-3]; Sedlack & Broadhurst, 1996), the maltreated children were predominantly from low-socioeconomic-status families. Consequently, demographically comparable nonmaltreated children were recruited from families receiving Temporary Assistance for Needy Families. A DHS recruitment liaison contacted eligible nonmaltreating families and described the project, and if interested, parents signed a release for their names to be given to the project for recruitment. DHS record searches were completed for these families to verify the absence of any record of child maltreatment. Trained research assistants also interviewed mothers of children recruited for the nonmaltreatment group to confirm a lack of DHS involvement and prior maltreatment experiences utilizing the Maternal Maltreatment Classification Interview (Cicchetti, Toth, & Manly, 2003). Subsequently, record searches were conducted in the year following camp attendance to verify that all available information had been accessed. Only children from families without any history of documented abuse or neglect were retained in the nonmaltreatment group. In addition, families who had received preventive services through DHS due to concerns over risk for maltreatment were excluded from the sample to reduce the potential for unidentified maltreatment existing within this group.

Classification into maltreatment subgroups was accomplished through use of the MCS (Barnett et al., 1993). The MCS is a reliable and valid method for classifying maltreatment (Bolger et al., 1998; English et al., 2005; Manly, Chicchetti, & Barnett, 1994) that utilizes DHS records detailing investigations and findings involving maltreatment in identified families over time. Rather than relying on official designations and case dispositions, the MCS codes all available information from DHS records, making independent determinations of maltreatment experiences. Based on operational criteria, the MCS designates all of the subtypes of maltreatment children have experienced (i.e., neglect, emotional maltreatment, physical abuse, sexual abuse) and the severity of each type on scales that range from 1 (*least*) to 5 (*most*) severe. Based on documented records, the MCS also determines when in the course of development maltreatment events occurred, providing indices of timing. Events were coded as occurring during five developmental periods, including infancy (0–12 months), toddlerhood (13–36 months), preschool (36–60 months), early school age (ages 5–7), and later school age (ages 8–12). The timing information allows for the determination of whether maltreatment occurred within

Table 1
Demographic Characteristics

	Early physical/ sexual abuse (n = 60)	Maltreated without early abuse (n = 205)	Nonmaltreated (n = 288)
	M (*SD*) or %	*M* (*SD*) or %	*M* (*SD*) or %
Age	9.91 (1.91)	10.02 (1.85)	10.08 (1.89)
Gender (% male)	70.0	58.5	44.4
Race/ethnicity			
African American	50.0	64.9	65.3
Latino	23.3	11.7	17.4
Caucasian	25.0	21.0	14.9
Other	1.7	2.4	2.4
Family history of welfare	96.7	95.0	96.0

each of the developmental periods. The developmental period of *onset* of maltreatment and the *recency* of maltreatment can be determined, and the number of developmental periods of maltreatment serves as an index of *chronicity*. Moreover, the developmental timing of each of the maltreatment subtypes also can be specified.

Coding of the DHS records was conducted by trained research assistants, doctoral students, and clinical psychologists. Coders were required to meet acceptable reliability with criterion standards before coding actual records for the study. Coders demonstrated acceptable reliability (weighted κs with the criterion ranging from .86 to .98). Reliabilities for the presence versus absence of maltreatment subtypes (κs = .90–1.00) and for individual subtypes severities (weighted κs = .75–1.00) also were established.

In terms of the subtypes of maltreatment, *neglect* involves failure to provide for the child's basic physical needs for adequate food, clothing, shelter, and medical treatment. In addition to inadequate attention to physical needs, forms of this subtype include lack of supervision, moral or legal neglect, and education neglect. *Emotional maltreatment* involves extreme thwarting of children's basic emotional needs for psychological safety and security, acceptance and self-esteem, and age-appropriate autonomy. Examples of emotional maltreatment of increasing severity include belittling and ridiculing the child, extreme negativity and hostility, exposure to severe marital violence, abandoning the child, and suicidal or homicidal threats. *Physical abuse* involves the nonaccidental infliction of physical injury on the child (e.g., bruises, welts, burns, choking, broken bones). Injuries range from minor and temporary to permanently disfiguring. Finally, *sexual abuse* involves attempted or actual sexual contact between the child and a family member or person caring for the child for purposes of that person's sexual satisfaction or financial benefit. Events range from exposure to pornography or adult sexual activity, to sexual touching and fondling, to forced intercourse with the child.

Children in the maltreatment group all had histories of abuse, neglect, or both occurring in their families according to DHS records. Among the maltreated children, 79.2% had experienced neglect (*M* neglect severity for those with neglect = 3.42, *SD* = 1.13), 67.8% had experienced emotional maltreatment (*M* severity = 3.44, *SD* = 1.24), 32.6% had experienced physical abuse (*M* severity = 2.56, *SD* = 1.04), and 8.3% had experienced sexual abuse (*M* severity = 2.85, *SD* = 0.94). The majority of children had experienced multiple subtypes of maltreatment. Specifically, 65.2% of the maltreated children had experienced two or more maltreatment subtypes, and 13 of the 15 possible combinations of presence versus absence of each of the four maltreatment subtypes were observed in the sample. On average, the maltreated children had experienced 1.89 (*SD* = 0.78) subtypes of maltreatment. The number of subtypes of maltreatment provides an index of maltreatment *diversity*.

Based on the MCS, we identified a group of children (*n* = 60) who had experienced *early abuse*, that is, physical or sexual abuse or both occurring prior to age 5 in the infancy, toddler, or preschool developmental periods (EPA/SA). In primary analyses, these children were contrasted with the remaining maltreated children (*n* = 205) who had *not* experienced early physical or sexual abuse (NEPA/SA), for example, experienced neglect or emotional maltreatment only or later onset of physical or sexual abuse.

Day Camp Procedures

Children attended a week-long day camp program and participated in research assessments (see Cicchetti & Manly, 1990, for detailed descriptions of camp procedures). Children were transported by bus to the camp each day, with travel time averaging 45 min. At the camp, children were assigned to groups of eight (four maltreatment, four control) same-age and same-sex peers. Each group was led by three trained camp counselors, who were unaware of the maltreatment status of children and the hypotheses of the study. Camp lasted 7 hr/day for 5 days, providing 35 hr of interaction between children and counselors. In addition to the recreational activities, after providing assent, children participated in various research assessments and provided morning and afternoon saliva samples. Trained research assistants, who also were unaware of research hypotheses and maltreatment status, conducted individual research sessions with children, in which questionnaires and other research measures were administered. Clinical consultation and intervention occurred if any concerns over danger to self or others emerged during research sessions. The counselors, who had been trained extensively for 2 weeks prior to the camp, also completed assessment measures on individual children, based on their 35 hr of observations and interactions with children in their respective groups. Counselors were unaware of the research hypotheses and children's maltreatment status.

Day Camp Measures

The camp context allowed for a setting where a consistent collection of saliva samples could occur at uniform times across the camp week. Each day upon arrival to the camp at 9 a.m., trained research assistants obtained saliva samples from each child. Given the approximate 45-min transportation time and the time spent being greeted by camp staff, children had been awake at least 1 hr prior to providing the morning saliva samples, thus avoiding the period of the dynamic cortisol awakening response (cf. Susman et al., 2007). Samples also were obtained each day prior to the children's departure from the camp at 4 p.m.

The children had not consumed food or drink for at least 30 min before each saliva sample was obtained. Samples were collected following the method recommended by Schwartz, Granger, Susman, Gunnar, and Laird (1999). The children were asked to chew Trident® sugarless original flavor gum to stimulate saliva flow and then passively drool although a short drinking straw into a 20-ml plastic vial. The samples were immediately frozen and stored at −40 °C. Each week, the samples were shipped overnight on dry ice for next day delivery to Salimetrics Laboratories (State College, PA) for assay. Cortisol was assayed in duplicate by Salimetrics Inc. using an enzyme immunoassay (Salimetrics Inc.). This assay has a lower limit of sensitivity of 0.007 μg/dl (range up to 1.8 μg/dl), and average intra- and interassay coefficients of variation less than 5.0% and 10.0%, respectively. Units of cortisol are expressed in μg/dl (micrograms per deciliter). Across camp days, the mean morning cortisol level was 0.19 (*SD* = 0.11), and the mean late afternoon level was 0.11 (*SD* = 0.07). As cortisol values were positively skewed and leptokurtotic, logarithmic transformations (log 10) of cortisol values were used for all statistical analyses. Values were averaged across days to yield one morning and one afternoon cortisol measure for each child.

Internalizing and Depressive Symptoms

Children's Depression Inventory (CDI). The CDI is a widely used self-report questionnaire to assess depressive symptomatology in school-age children (Kovacs, 1982, 2004). For each item, children chose from among three option statements, depicting increasing levels of depressive symptoms, in order to characterize their experiences in the past 2 weeks. Kovacs (2004) reports that internal consistency for the total scale has ranged from 0.71 to 0.89, and validity has been well established. In the current investigation, internal consistency was 0.85. A total score of 19 and above has been established as a cutpoint for clinical level depressive symptoms (Kovacs, 2004). In the current sample, 11.4% ($n = 63$) of the children reported clinical-level depressive symptoms.

Teacher Report Form (TRF). Behavioral symptomatology was evaluated at the end of each week by counselors' completion of the TRF (Achenbach, 1991). The TRF is a widely used and validated instrument to assess behavioral disturbance from the perspective of teachers, and the measure was used in the present study because camp counselors are able to observe similar behaviors to that of teachers. The TRF, containing 118 items rated for frequency, assesses two broadband dimensions of child symptomatology, internalizing and externalizing, as well as total behavior problems. In the present investigation, we focus on the internalizing dimension, which includes subscales of withdrawn, somatic problems, and anxiety/depression. Interrater reliability for the internalizing scale was based on average intraclass correlations among pairs of raters. Across the years, reliabilities for internalizing ranged from 0.56 to 0.84 ($M = 0.68$), consistent with the greater difficulty in observers rating internalizing symptoms (Achenbach, 1991). Internal consistency was 0.84. The counselors' scores for each child were averaged to obtain individual child scores for the internalizing dimension. Utilizing the *T* scores for the internalizing scale, a cutpoint of 60T was used to designate the presence or absence of high internalizing symptoms (Achenbach, 1991). In the sample, 9.9% of the participants ($n = 55$) were classified as having high-level internalizing problems.

Summary measure of internalizing. To yield a multi-informant index of children displaying high depressive and internalizing symptomatology, child and counselor measures were combined. Overall, the CDI and TRF internalizing scores were significantly but not strongly correlated, $r = .20$, $p < .001$. Similarly, the high symptom level groups for the CDI and TRF were significantly but weakly associated, $\chi^2(1, N = 553) = 4.48$, $p = .03$. This level of correspondence is consistent with the degree of agreement typically found between different sources of report for child internalizing symptoms. To combine both perspectives, the presence of high symptoms on either the CDI or the TRF was determined and used as an index of high depressive and internalizing symptomatology, resulting in 19.3% of the children being classified as symptomatic.

Results

Plan of Analysis

The analysis was complicated by the complexity of maltreatment subgroup categorization, based on maltreatment subtype and developmental timing of individual subtypes, and concomitant cell sizes, particularly when participants were further classified by the presence of high versus low depressive or internalizing symptoms. We, therefore, pursued a layered approach to these analyses. First, maltreatment parameters of the EPA/SA were examined relative to characteristics of other maltreated children (NEPA/SA) for descriptive purposes. Second, we examined whether the EPA/SA group differed from the NEPA/SA and NC children on depressive and internalizing symptoms, as well as morning and afternoon cortisol levels. Third, we examined main and interaction effects of early abuse and depressive or internalizing symptoms on cortisol levels and change from morning to late afternoon. Gender and age effects also were probed. Fourth, using only the maltreated children, we examined the specificity of EPA/SA: first, relative to *early* neglect or emotional maltreatment without early abuse, and next, relative to *later onset* of physical or sexual abuse at age 5 or older. Finally, other aspects of maltreatment timing, diversity, and perpetrator were considered to determine whether these features of maltreatment influenced cortisol regulation in the context of high symptomatology.

Characteristics of the EPA/SA Group

Within the EPA/SA group, 80% had experienced physical abuse and 28.3% had experienced sexual abuse; 5 children had experienced both maltreatment subtypes. Physical and sexual abuse often occurred in the presence of neglect and emotional maltreatment. In fact, 73.3% of the children in the EPA/SA group had experienced neglect, and 73.3% had also experienced emotional maltreatment. Overall, the EPA/SA group had been exposed to an average of 2.55 subtypes of maltreatment (*SD* = 0.70). The individual subtype severities for children who had experienced respective subtypes did not differ between the EPA/SA and NEPA/SA groups (*ps* = .15 to .77).

Among the maltreated children, 89.8% had experienced some form of maltreatment from their mothers, and this rate did not differ for the EPA/SA (91.7%) versus NEPA/SA children (89.3%) groups, $\chi^2(1, N = 265) = 0.29$, $p = .81$. The rate of fathers as perpetrator also was comparable by group, $\chi^2(1, N = 265) = 0.04$, $p = .88$, 43.3% versus 42.0%. However, the rate of others as perpetrator was higher in the EPA/SA group (58.3%) than the NEPA/SA group (35.1%), $\chi^2(1, N = 265) = 10.39$, $p = .002$.

For the EPA/SA group, the onset of maltreatment had occurred in infancy for 43.3%, in toddlerhood for 25.0%, and in preschool for 31.7% of the early abuse group. Moreover, 63.3% of the children had been maltreated during the toddler period and 93.3% had been maltreated in the preschool period. However, in terms of recency, only 18.4% had been maltreated beyond the preschool period. Chronicity of maltreatment, as indexed by the number of developmental periods of maltreatment, averaged 2.22 developmental periods (*SD* = 0.92).

Group Differences on Depressive and Internalizing Symptoms and Cortisol

Analyses of variance were conducted to contrast children in the EPA/SA, NEPA/SA, and NC groups on CDI and TRF internalizing scores (see Table 2). Results differed depending on whether child self-report or counselor reports were used. For child self-reported symptoms on the CDI, the main effect of group was marginally significant, $F(2, 550) = 2.62$, $p = .07$. In contrast, significant differences for group classification were found for counselor-reported TRF internalizing symptoms, $F(2, 550) = 9.14$, $p < .001$. Tukey post hoc tests indicated that the EPA/SA group had higher internalizing scores than the NEPA/SA and NC children.

As described in the Method section, the presence of high symptoms on the CDI or the TRF was used as an index of high depressive and internalizing symptomatology. Significant group differences were observed on the presence or absence of high symptoms, $\chi^2(2, N = 553) = 11.71$, $p = .003$. Follow-up contrasts indicated that the EPA/SA group (33.3%) had significantly higher rates of symptoms than the NC children (14.9%), $\chi^2(1, N = 348) = 11.34$, $p = .001$, and marginally higher rates than the NEPA/SA children (21.5%), $\chi^2(2, N = 265) = 3.57$, $p = .06$. Finally, the NEPA/SA maltreated children had marginally higher rates of ele-

Table 2

Comparison of Early Physical/Sexual Abuse Classification Groups on Depressive and Internalizing Symptoms and Cortisol

	Early physical/sexual abuse (n = 60)	Maltreated without early abuse (n = 205)	Nonmaltreated (n = 288)
	M (*SD*)	*M* (*SD*)	*M* (*SD*)
CDI total score	10.43 (7.84)	9.32 (7.75)	8.30 (6.70)
TRF internalizing *T* score	51.79 (9.48)$_a$	48.91 (7.99)$_b$	47.29 (7.15)$_b$
	% (*n*)	% (*n*)	% (*n*)
High CDI or TRF	33.3 (20)$_a$	21.5 (44)$_{ab}$	14.9 (43)$_b$
	M (*SD*)	*M* (*SD*)	*M* (*SD*)
Average morning cortisol	−0.82 (0.19)	−0.79 (0.20)	−0.78 (0.18)
Average afternoon cortisol	−1.04 (0.16)	−1.03 (0.17)	−1.03 (0.15)

Note. Groups having the same subscript are not significantly different. Cortisol values are log transformed (log 10). CDI = Children's Depression Inventory; TRF = Teacher's Report Form.

vated symptoms than the nonmaltreated children, $\chi^2(2, N = 493) = 3.52$, $p = .06$. Finally, as shown in Table 2, the groups did not differ on average morning, $F(2, 550) = 1.52$, $p = .22$, or afternoon cortisol levels, $F(2, 550) = 0.07$, $p = .94$.

Diurnal Cortisol Regulation: The Interaction of EPA/SA With Depressive and Internalizing Symptomatology

A 3 (EPA/SA groups) × 2 (depression/internalizing level) × 2 (AM/PM cortisol) repeated measures analysis of covariance (ANCOVA) was conducted, covarying age and gender (see Table 3). In addition to significant gender and age effects, a significant main effect was found for symptom level and a marginally significant main effect was obtained for EPA/SA classification group. However, these main effects were qualified by a significant interaction among group, symptom level, and AM/PM, $F(2, 545) = 6.07$, $p = .002$. Follow-up tests for simple effects were conducted separately within the high and low symptom levels. As shown in Figure 1a, for children who were not high on depressive or internalizing symptoms, no effect of group on diurnal variation in cortisol was obtained, $F(2, 441) = 0.81$, $p = .45$. As shown in Figure 1b, for children with high depressive or internalizing problems, group status was a significant predictor of diurnal cortisol activity, $F(2, 104) = 5.94$, $p = .004$. To further pinpoint the effect for the high-symptom children, we computed the AM to PM decline (AM minus PM) and subjected these scores to an ANCOVA now entering morning cortisol, gender, and age as covariates. The effect of group was significant, $F(2, 101) = 4.27$, $p = .017$. Bonferroni-corrected post hoc comparisons indicated that the EPA/SA group had a significantly lower diurnal decline in cortisol (0.12) than the NEPA/SA group (0.23) and the NC group (0.25). The morning and afternoon scores also were examined separately. Although the group differences were in the expected direction (see Figure 1b), the groups did not differ significantly at either time of day. Thus, the EPA/SA group with high depressive or internalizing symptoms was distinguished by the smaller decrease or attenuated slope of the daytime diurnal rhythm of cortisol, due to slightly but not significantly lower AM and slightly but not significantly higher afternoon levels than the other children.

Table 3

Repeated Measures (Morning and Afternoon Cortisol) Analysis of Covariance for High Depressive or Internalizing Symptoms and Early Physical or Sexual Abuse Classification Groups

Source	*df*	*F*	η^2	*p*
Within subjects				
Time	1	0.66	.001	.42
Sex × Time	1	5.87	.011	.016
Age × Time	1	28.28	.049	.000
Early Physical/Sexual Abuse × Time	2	2.36	.009	.095
High Symptoms × Time	1	6.46	.012	.011
Early Physical/Sexual Abuse × High Symptoms × Time	2	6.07	.022	.002
Error	545			

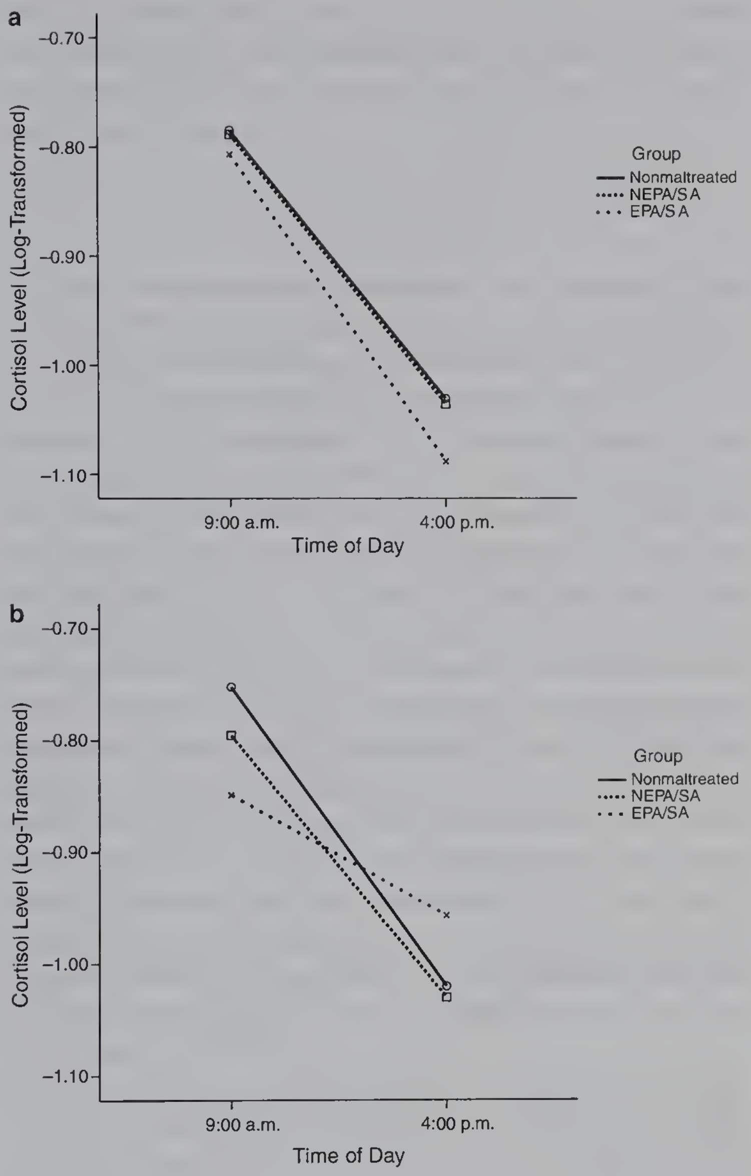


Figure 1. (a) Children with *low* depressive and internalizing symptoms: Diurnal cortisol regulation in early abuse classification groups. (b) Children with high depressive or internalizing symptoms: Diurnal cortisol regulation in early abuse classification groups.

Note. Cortisol values are in μg/dl with log 10 transformation, adjusted for covariates. EPA/SA = early physical or sexual abuse; NEPA/SA = maltreated, but not early physical or sexual abuse.

The effects of the covariates of gender and age on the within subjects diurnal decrease in cortisol are represented in the repeated measures ANCOVA by the respective interactions of the covariates with time. To interpret the effects of the covariates, the difference between morning and afternoon cortisol was computed, and t tests were conducted for gender and age groups (< 10 and ≥ 10). Gender differences were found, $t(551) = 2.83$, $p = .005$, such that girls had a significantly greater cortisol decrease across the day ($M = 0.27$, $SD = 0.17$) than boys ($M = 0.22$, $SD = 0.19$). Although morning cortisol levels did not differ by gender, $t(551) = 0.74$, $p = .46$, afternoon levels did, $t(551) = 2.41$, $p = .02$, with boys having higher cortisol ($M = -1.02$, $SD = 0.16$) than girls ($M = -1.05$, $SD = 0.16$). Interpretation of the covariate effect of age involved a significant effect of age group on diurnal cortisol decrease, $t(551) = 6.12$, $p < .001$, with older children evincing a steeper decrease ($M = 0.29$, $SD = 0.19$) than younger children ($M = 0.19$, $SD = 0.17$). Older children had significantly higher morning cortisol levels than younger children, $t(550.9) = 5.46$, $p < .001$ ($M_{\text{older}} = -0.74$, $SD = 0.20$; $M_{\text{younger}} = -0.83$, $SD = 0.17$). Afternoon cortisol did not differ, $t(551) = 0.56$, $p = .58$.

Because puberty is associated with a rise in basal and reactivity measures of the HPA axis (Adam, 2006; Gunnar, Wewerka, Frenn, Long, & Griggs, 2009; Stroud et al., 2009) and is highly correlated with age, and because age was a significant covariate in the above analyses, we examined whether age might interact with the Group × Symptom Level effects on the diurnal decline in cortisol reported above. To examine this question, we used a regression analysis with diurnal decrease as the dependent factor. EPA/SA versus others, age group, and symptom level were entered as dummy variables, followed two-way interaction terms, and finally the three-way interaction of age, symptom level, and group. Importantly, the results did not yield a significant three-way interaction, $\beta = .01$, $p = .73$, indicating that age did not moderate the overall findings observed involving attenuated diurnal decrease in children with early abuse and high symptoms. Similarly, we considered whether gender would moderate the effects of early abuse and symptom level. An analogous regression analysis, as was conducted for age, also was carried out with gender; however, the three-way interaction of early abuse, symptom level, and gender also was not significant, $\beta = -.01$, $p = .82$, indicating that the effects were not operating differently among boys and girls.

EPA/SA Versus Early Neglect/Emotional Maltreatment

To examine whether other forms of early maltreatment exerted an impact on the diurnal decrease in cortisol, we examined only the children whose maltreatment onset was prior to age 5. There were 151 children who experienced neglect or emotional maltreatment but not physical or sexual abuse, and these children were contrasted with the EPA/SA group. The severity of emotional maltreatment and of neglect did not differ between these two groups. In an ANCOVA using the decrease in cortisol as the dependent measure, the early maltreatment subtype designation and high versus low symptom status were used as grouping factors, and morning level cortisol, age, and gender as covariates. A significant main effect was found for symptom status, $F(1, 204) = 13.13$, $p < .001$; however, this effect was clarified by a significant interaction effect of symptom status and maltreatment type, $F(1, 204) = 7.29$, $p = .007$. To probe this interaction effect, the four groups in the interaction were contrasted with Bonferroni post hoc tests. The interaction effect is shown in Figure 2a. The EPA/SA group with high symptoms ($M = 0.13$, $SD = 0.23$) showed a smaller diurnal decrease than any of the other groups. The difference was highly significant for each of the low depressive symptom groups, early neglect-emotional maltreatment ($M = 0.24$, $SD = 0.17$, $p = .004$) and EPA/SA ($M = 0.27$, $SD = 0.17$, $p = .001$), and was marginally significant for the early neglect, high depressive symptom group ($M = 0.22$, $SD = 0.17$, $p = .10$). No other contrast was significant.

The Effect of Early Versus Late Onset of Physical/Sexual Abuse

We also sought to determine whether EPA/SA was responsible for the observed effects, or whether physical or sexual abuse with later onset also would demonstrate the attenuated diurnal decrease in cortisol. Accordingly, children who experienced the onset of physical or sexual abuse beyond the preschool period were identified ($n = 41$), and these children were contrasted with the EPA/SA group. The severity of physical abuse and of sexual abuse did not differ between these abuse onset groups. In an ANCOVA using the decrease in cortisol as the dependent measure, the abuse onset group and high versus low symptom status were used as grouping factors, and morning level cortisol, age, and gender as covariates. As shown in Figure 2b, a significant interaction between timing of abuse onset (< 60 months, ≥ 60 months) and the presence of high depressive or internalizing symptoms was found, $F(1, 94) = 11.87$, $p = .001$. The interaction was probed through Bonferroni-corrected post hoc comparisons. The EPA/SA group with high symptoms ($M = 0.12$, $SD = 0.24$) had a significantly lower decrease in cortisol than the later abuse/low level symptom group ($M = 0.29$, $SD = 0.20$, $p = .05$) and the EPA/SA group with low level symptoms ($M = 0.27$, $SD = 0.17$, $p = .001$); however, a significant contrast was not found with the later abuse and low level symptom group ($M = 0.21$, $SD = 0.20$, $p = .25$). Other contrasts were not significant. These results again highlight the EPA/SA—high symptom group as uniquely demonstrating a pattern of attenuated diurnal decrease in cortisol.

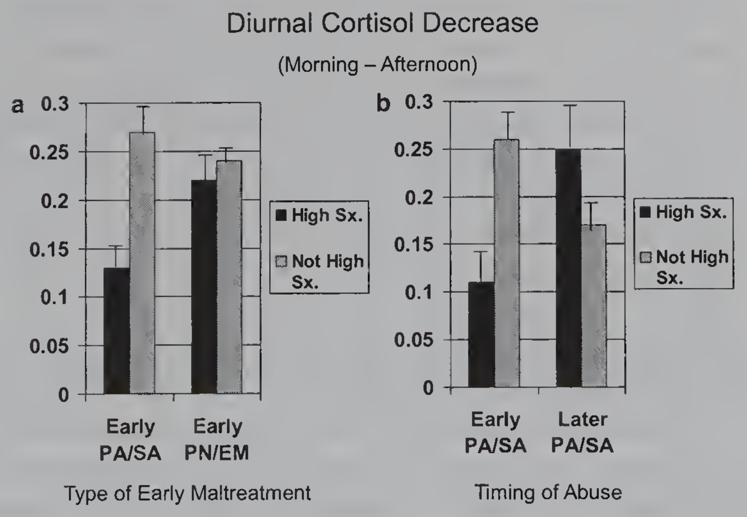


Figure 2. Specificity of diurnal cortisol attenuation for early physical or sexual abuse and high symptoms: Comparisons with early neglect or emotional maltreatment and with later physical or sexual abuse.

Note. Diurnal cortisol decrease values are adjusted for covariates. PA = physical abuse; SA = sexual abuse; PN = physical neglect; EM = emotional maltreatment; Sx. = internalizing and depression symptoms.

Features of Maltreatment Timing, Diversity, and Perpetrator

Finally, we considered whether the onset period of maltreatment, recency of maltreatment, chronicity of maltreatment (number of developmental periods of maltreatment), diversity of maltreatment experiences (number of subtypes of maltreatment), and mother, father, and others as perpetrators of maltreatment provided more discriminating perspectives on the atypical pattern of cortisol change across the day observed in early physically or sexually abused/high symptom children. For each of these seven maltreatment features, analogous repeated measures ANCOVAs were conducted, as had been done with the early abuse group classification. High versus low depressive and internalizing symptom status was included as a second main effect, with age and gender as covariates. We determined if any significant interaction effects for time of day, symptom status, and the respective maltreatment feature were significant in relation to cortisol regulation across the day. None of the maltreatment parameters we considered resulted in a significant interaction effect (*ps* = .26, .46, .32, .19, .48, .16, and .99, for onset, chronicity, recency, diversity, mother as perpetrator, father as perpetrator, and other as perpetrator of maltreatment, respectively).

Discussion

In this investigation we sought to examine whether children experiencing physical or sexual abuse in the early years of life and exhibiting high internalizing symptomatology displayed a unique signature of cortisol levels across the day suggestive of neuroendocrine dysregulation. The results supported our predictions. First, we found that children experiencing abuse in the first 5 years of life exhibited more internalizing symptoms than maltreated children without early abuse and nonmaltreated children of the same socioeconomic level. Second, children experiencing early abuse who had high internalizing symptoms exhibited an atypical flattening of cortisol production over the daytime hours. Further analyses indicated that even among children high on internalizing symptoms, other parameters of maltreatment, including developmental period of onset, recency, chronicity, diversity, and type of perpetrator, did not provide a better explanation of the phenomena than did the experience of EPA/SA.

The pattern of cortisol production noted for the early abused/high internalizing children provided consistent evidence of HPA axis dysregulation as noted in adults and children under chronic stress (Gunnar & Vazquez, 2001; Heim et al., 2008; Miller et al., 2007). That is, we obtained evidence of a less marked decline in cortisol over the daytime due to a slight lowering of AM and slight increase in PM levels. Given the times we were able to sample saliva, it is somewhat remarkable that we were able to detect this pattern of dysregulation. Cortisol is at its daily peak approximately 30 min after morning awakening, drops steeply for the next 30 min to 1 hr, and then declines gradually over the day, until it reaches almost a zero nadir soon after the onset of nighttime sleep (cf. Susman et al., 2007). Most studies that have identified this stress-related pattern of daytime cortisol dysregulation have sampled at morning wake-up and at bedtime. In the present study, in contrast, while we sampled the children at 9 a.m. it is very likely that the children had been awake long enough so that we missed the early morning peak. Although 4 p.m. is late in the afternoon, levels at this time are not as low as the nadir that they will reach later in the evening. Thus, it is possible that the pattern we noted would have been even more marked had we been able to sample earlier and later in the day.

Whereas the pattern we noted is not the same as that seen among adults with clinical depression, the findings are consistent with those described in two other reports on depressed, maltreated children studied in a day camp setting (Hart et al., 1996; Kaufman, 1991). Hart et al., (1996) found that maltreated children who were also depressed showed a less marked diurnal decrease in cortisol from 9 a.m. to 4 p.m., although no attempt was made in that study to differentiate patterns based on subtype of maltreatment. Kaufman (1991) noted that maltreated and depressed children in her sample were less likely than nondepressed maltreated children to show a decrease in cortisol from 9 a.m. to 4 p.m. Notably similar to the present study, she also observed that the depressed maltreated children were more likely to have suffered abuse in addition to neglect, rather than solely neglect. Neither the Hart et al. study nor the Kaufman study examined the influence of the developmental timing of maltreatment subtypes.

We also found that both gender and age were associated with morning to afternoon change in cortisol level. Whereas girls had significantly lower afternoon cortisol and greater diurnal decrease than boys, gender did not moderate the effects of EPA/SA and high symptoms. In terms of age, consistent with the literature, older children were found to have significantly higher morning cortisol levels and a greater morning to afternoon cortisol decrease than younger children. However, age did not moderate the overall effects of EPA/SA and high symptoms. Thus, the findings are generally applicable for boys and girls, as well as for children across the age range studied.

The results of this investigation also contribute to our knowledge of pathways to depressive and internalizing symptomatology. For example, Kaplow and Widom (2007) reported that an earlier occurrence of maltreatment, defined as onset of maltreatment in the first 5 years of life, predicted more symptoms of anxiety and depression in adulthood. Our findings suggest that the results of Kaplow and Widom may have been even more compelling if they had been able to examine physical abuse, sexual abuse, and neglect separately. However, both the Kaplow and Widom and the current investigation raise the question as to why an early onset of maltreatment, and in our case abuse, is more predictive than a later onset in increasing risk for internalizing and depressive symptomatology and, in the context of these symptoms, dysregulated patterns of diurnal cortisol production. Early abuse may be more damaging to developing emotion and stress systems because it occurs during periods of rapid neurodevelopment (Cicchetti & Walker, 2001; Gunnar & Vazquez, 2006). Very young children also may be less able than older children to discern the cues predictive of an abusive attack, and this lack of predictability may engender chronic stress and hypervigilance to aggression in these youngsters, even when abusive events are not occurring (Rieder & Cicchetti, 1989; Teisl & Cicchetti, 2008). School-aged physically abused children, on the other hand, have been shown to be highly sensitive to even very degraded signals of threat, which although potentially maladaptive in some contexts, may provide some relief from constant vigilance in the abusive home (Pollak, Cicchetti, & Klorman, 1998; Pollak & Sinha, 2002; Shackman, Shackman, & Pollak, 2007). In addition, the impact of EPA/SA on developing brain systems may be especially pernicious because it occurs during a period when the child is nearly wholly dependent on parents for survival. For the abused infant, toddler, or preschooler, a hypervigilant state of mind and chronic stress with respect to unpredictable parental attacks may shift neurobiological development onto pathways leading to depression and neuroendocrine dysregulation (Cicchetti & Rogosch, 2001; Tarullo & Gunnar, 2006).

Another question pertains to why early neglect in the absence of physical or sexual abuse did not have a significant impact on either internalizing and depressive symptoms or HPA axis activity. While early neglect has been shown to have marked impacts on young children's social and cognitive competence (Pears & Fisher, 2005), neglect experiences may not be as fear and threat arousing as are the experiences of being physically attacked and hurt. Although it is not uncommon to find early neglect in the absence of abuse, it is rare to find early abuse in the absence of neglect, especially in low-income samples such as the one we studied. Nearly three fourths of our EPA/SA sample also was neglected and emotionally maltreated. We do not know whether their neglect and emotional maltreatment contributed to their increased internalizing symptomatology and HPA dysregulation; however, it is possible that neglect made these children even more vulnerable to the fear and stress engendered by early abuse.

It is also conceivable that our findings may have been even more striking if we had included maltreated children who were residing in foster care. Generally children who experience early and extreme maltreatment are removed from their homes and placed in foster care. Because we excluded children with out-of-home placements, it is likely that we did not include the most extreme cases. This is not to say that the maltreated children in our sample experienced only mild abuse. Indeed, the median physical or sexual abuse severity score as coded using the MCS for the early abused children was 3 on the scale of 5 points, corresponding to incidents involving the infliction of marks on a child's face, serious bruises, or fondling a child for sexual gratification. Rather, by excluding children with out-of-home placements, we avoided confounding the impact of maltreatment with the stress of transitions in and out of foster care homes. These changes may be important in the sequelae of emotion and neuroendocrine dysregulation for many maltreated children (Gunnar & Fisher, 2006; Kaufman, 1991). Studying preschool children who had primarily experienced neglect, Bruce, Fisher, Pears, and Levine (in press) found reduced morning cortisol in the first months following a new foster placement. A year or more later, however, whether or not the morning cortisol recovered, continued to be dysregulated, or became more dysregulated depended on whether the children were randomly assigned to regular foster care or to a foster care intervention condition that stabilized their early care (Fisher, Stoolmiller, & Gunnar, 2007). Thus, the Bruce et al. study revealed the sensitivity of cortisol regulation to transient impacts of neglect and relationship disruption, while the present study indicated the long-term impacts of EPA/SA in the absence of the child being removed from the home.

Consistent with the results of Heim et al. (2008), the combination of high depressive symptoms and EPA/SA appeared to be critical to dysregulation of the HPA axis. Early physically and sexually abused children who were low on depressive and internalizing symptoms had cortisol patterns that were comparable to other maltreated and nonmaltreated children. However, whereas Heim et al. examined reactions of the HPA axis to psychosocial stressors and pharmacological challenges, we addressed only the children's typical daytime pattern of cortisol production. We do not know whether the high internalizing, early abused children also would show larger or more prolonged cortisol responses to a stressor task similar to the adult women in the Heim et al. study. This awaits further research. It also will be important to conduct longitudinal studies in order to ascertain whether the dysregulated pattern of cortisol production noted among these high internalizing/early abused children presages ongoing patterns of neurobiological and neuroendocrine dysregulation as the children develop.

There are many parallels between the neurobiological effects of early abuse and those of MDD (Heim et al., 2004; Heim et al., 2008). It is likely that prior investigations of individuals with MDD have included persons who had experienced early abuse (Bemporad & Romano, 1992; Heim et al., 2008). As such, many of the established neurobiological findings in depression may be due to the undetected presence of EPA/SA (Heim et al., 2004). In the absence of early abuse, depression may not be related to changes in the HPA system. Thus, there may be different neurobiological subtypes of depression. Accordingly, a history of early abuse, as well as the clinical presentation, onset, and course of depression should be incorporated into the formulation of potential etiological models of mood disorder (Heim et al., 2004; Heim et al., 2008). For example, because cortisol dysregulation rarely occurs in depressed children who did not experience abuse early in life (Feder et al., 2004), childhood depression without EPA/SA may be caused by different processes and have different neurobiological features than childhood depression with early abuse. Our findings that nonmaltreated children with high levels of depressive or internalizing symptomatology did not exhibit atypical cortisol regulation attest to this divergence. There may be biologically distinct subtypes of depression as a function of the presence or absence of abuse early in life (cf. Heim et al., 2004; Heim et al., 2008). Thus, not all individuals with depression exhibited cortisol dysregulation; we found it only for the early physically or sexually abused children with high internalizing and depressive symptomatology—a group that is more likely to have dysfunctions in early neurobiological development through the high allostatic load caused by abuse.

It is also important to note that not all of the children who experienced early abuse had high depressive or internalizing symptoms. Why is the HPA axis functioning of this group of early abused children better regulated than that of early abused and depressed children with the attenuated diurnal decrease in cortisol? It is likely that there are genes that increase the probability of depression and internalizing problems under conditions of high stress such as EPA/SA. For example, a number of investigations have demonstrated that the s/s and s/l genotypes of the serotonin transporter gene linked promoter region (5-HTTLPR), in interaction with severe child maltreatment, are associated with increased risk for depression in children, adolescents, and adults (see, e.g., Caspi et al., 2003; Cicchetti, Rogosch, & Sturge-Apple, 2007; Kaufman, Cook, Amy, Jones, & Pittinsky, 2004). Conversely, the l/l genotype of the 5-HTTLPR may confer a protective function by decreasing the probability that individuals experiencing severe maltreatment will develop depression (Cicchetti et al., 2007; Kaufman et al., 2004). Consistent with calls for a multiple levels of analysis approach to investigating the determinants of resilience (Charney, 2004; Cicchetti & Curtis, 2007; Curtis & Cicchetti, 2003), future research will need to include assessments of potential genetic dispositions to depression to discover whether individuals experiencing EPA/SA may have the genes that confer vulnerability or protection against depression (see also Bradley et al., 2008). Similarly, it will be important in subsequent investigations to examine the contribution that the CRH genes and the glucocorticoid receptor genes play in individual differences in HPA axis activity, regulation, and reactivity.

There are several limitations in the design that deserve discussion. First, it is impossible to isolate completely the effects of physical and sexual abuse from those of neglect. The clinical reality is that there is an extremely high overlap and that neglect is almost a consistent experience for these children, particularly in low-income samples. Thus, some caution is warranted in interpreting the meaning of our subtype findings. It may be, as we argued above, that abuse (physical or sexual) has features (e.g., provoke feelings of fear) that are particularly potent in activating stress and emotion pathways and thus in dysregulating rapidly developing stress and emotion neurobiological systems. However, it could also be that abuse in the context of neglect is a more severe form of maltreatment and thus that "severity" rather than particular features (e.g., fear provoking) produced the effects we observe. As abuse and neglect are nearly impossible to disentangle in human populations, it may prove useful to translate the findings from the current study back into animal models that compare early abuse, early abuse and early neglect, early neglect only, and later abuse and neglect.

Another limitation in the study was our lack of information about the time when the children awoke. Cortisol levels change markedly in the 1st hour after morning awakening and then gradually decrease over the remainder of the day. Levels are more closely related to the individual's wake-up time than to clock time. We chose not to measure wake-up time because we anticipated errors in child reporting of this measure, given the high level of disorganization in many of the families' homes. However, in future studies determining a valid method of obtaining individual wake-up times would be helpful. Note, however, that in order for differences in wake-up time to have explained our results, high symptom children with physical or sexual abuse prior to age 5 would have to have woken up earlier than all the other children. Although it is possible, it seems unlikely that this was the case.

A third limitation was that we did not obtain puberty-stage data on these children. Puberty is associated with increases in cortisol levels and reactivity to stressors (Gunnar et al., 2009; Stroud et al., 2009). Undoubtedly some of the older children in our study were entering puberty and this may have influenced their cortisol levels. In the absence of pubertal measures, we used age group (under age 10, age 10, and older) as a proxy measure, and as reported earlier, we did not find that age level moderated the effects of early abuse and high symptoms. Thus, we do not have evidence that pubertal development in the current sample had a major impact on the findings.

Finally, we cannot ascertain whether differences in the diurnal decrease in cortisol of the magnitude we observed will be predictive of future physical or mental health risks in children. There is increasing interest in the impact of chronic stress on patterns of diurnal HPA axis activity. Several studies of adult cohorts now show that "flattening" of the diurnal decrease in the same range as we observed predicts heightened risk of impairment in cardiovascular and immune system functioning (Heim et al., 2008; Raison & Miller, 2003). Thus it is conceivable that variations in the diurnal decrease in the range we report reflect mechanisms that have health consequences. Nonetheless, we cannot conclude that this is the case without longitudinal studies.

Our findings possess significant implications for children within the child welfare population, particularly with respect to the importance of early preventive interventions for children who have been abused or neglected even in the absence of clinical levels of depressive or internalizing symptomatology. In addition, intervention, timed as closely as possible to the early abusive experience, could be implemented to prevent the recidivism of maltreatment, to redirect the abused children onto a positive developmental trajectory, and to prevent the emergence of depression (Cicchetti, Rogosch, & Toth, 2006; Toth & Cicchetti, 1993). Because HPA axis dysregulation in early abused children with depression may reflect a unique subtype of depression, differential responses to treatment modalities may emerge, and this is an area requiring further investigation. Given what we know about the effects of cortisol dysregulation on a wide array of neurobiological systems, as well as on physical health and cognitive processes, the results of this investigation also have significant public health implications.

References

Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist/4-18 and 1991 profile*. Burlington: University of Vermont, Department of Psychiatry.

Adam, E. K. (2006). Transactions among adolescent trait and state emotion and diurnal and momentary cortisol activity in naturalistic settings. *Psychoneuroendocrinology, 31*, 664–679.

Barnett, D., Manly, J. T., & Cicchetti, D. (1993). Defining child maltreatment: The interface between policy and research. In D. Cicchetti & S. L. Toth (Eds.), *Child abuse, child development, and social policy* (pp. 7–74). Norwood, NJ: Ablex.

Barr, C. S., Newman, T. K., Shannon, C., Parker, C., Dvoskin, R. L., Becker, M. L., et al. (2004). Rearing condition and rh5-HTTLPR interact to influence limbic-hypothalamic-pituitary adrenal axis response to stress in infant macaques. *Biological Psychiatry, 55*, 733–738.

Bemporad, J. R., & Romano, S. J. (1992). Childhood maltreatment and adult depression: A review of research. In D. Cicchetti & S. L. Toth (Eds.), *Rochester symposium on developmental psychopathology: Vol. 4. Developmental perspectives on depression* (pp. 351–376). Rochester, NY: University of Rochester Press.

Bolger, K. E., Patterson, C. J., & Kupersmidt, J. B. (1998). Peer relationships and self-esteem among children who have been maltreated. *Child Development, 69*, 1171–1197.

Bradley, R. G., Binder, E. B., Epstein, M. P., Tang, Y., Nair, H. P., Liu, W., et al. (2008). Influence of child abuse on adult depression: Moderation by the corticotropin-releasing hormone receptor gene. *Archives of General Psychiatry, 65*, 190–200.

Bruce, J., Fisher, P. A., Pears, K. C., & Levine, S. (in press). Morning cortisol levels in preschool-aged foster children: Differential effects of maltreatment type. *Developmental Psychobiology*.

Carrion, V. G., Weems, C. F., Ray, R. D., Glaser, B., Hessl, D., & Reiss, A. L. (2002). Diurnal salivary cortisol in pediatric posttraumatic stress disorder. *Biological Psychiatry, 51*, 575–582.

Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H. L., et al. (2003). Influence of life stress on depression: Moderation by a polymorphism in the 5-HTT gene. *Science, 301*, 386–389.

Charney, D. (2004). Psychobiological mechanisms of resilience and vulnerability: Implications for successful adaptation to extreme stress. *American Journal of Psychiatry, 161*, 195–216.

Cicchetti, D., & Barnett, D. (1991). Toward the development of a scientific nosology of child maltreatment. In W. Grove & D. Cicchetti (Eds.), *Thinking clearly about psychology: Essays in honor of Paul E. Meehl: Personality and psychopathology* (Vol. 2, pp. 346–377). Minneapolis: University of Minnesota Press.

Cicchetti, D., & Curtis, W. J. (Eds.). (2007). A multi-level approach to resilience [Special issue]. *Development and Psychopathology, 19*(3).

Cicchetti, D., & Manly, J. T. (1990). A personal perspective on conducting research with maltreating families: Problems and solutions. In G. Brody & I. Sigel (Eds.), *Methods of family research: Families at risk* (Vol. 2, pp. 87–133). Hillsdale, NJ: Erlbaum.

Cicchetti, D., & Rizley, R. (1981). Developmental perspectives on the etiology, intergenerational transmission and sequelae of child maltreatment. *New Directions for Child Development, 11*, 31–55.

Cicchetti, D., & Rogosch, F. A. (1996). Equifinality and multifinality in developmental psychopathology. *Development and Psychopathology, 8*, 597–600.

Cicchetti, D., & Rogosch, F. A. (2001). Diverse patterns of neuroendocrine activity in maltreated children. *Development and Psychopathology, 13*, 677–694.

Cicchetti, D., & Rogosch, F. A. (2001). The impact of child maltreatment and psychopathology upon neuroendocrine functioning. *Development and Psychopathology, 13*, 783–804.

Cicchetti, D., Rogosch, F. A., & Sturge-Apple, M. L. (2007). Interactions of child maltreatment and 5-HTT and monoamine oxidase A polymorphisms: Depressive symptomatology among adolescents from low-socioeconomic status backgrounds. *Development and Psychopathology, 19*, 1161–1180.

Cicchetti, D., Rogosch, F. A., & Toth, S. L. (2006). Fostering secure attachment in infants in maltreating families through preventive interventions. *Development and Psychopathology, 18*, 623–650.

Cicchetti, D., Toth, S. L., & Manly, J. T. (2003). *Maternal Maltreatment Classification Interview.* Unpublished manuscript, Mt. Hope Family Center, Rochester, NY.

Cicchetti, D., & Valentino, K. (2006). An ecological transactional perspective on child maltreatment: Failure of the average expectable environment and its influence upon child development. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology: Vol. 3. Risk, disorder, and adaptation* (2nd ed., pp. 129–201). New York: Wiley.

Cicchetti, D., & Walker, E. F. (2001). Stress and development: Biological and psychological consequences. *Development and Psychopathology, 13*, 413–418.

Curtis, W. J., & Cicchetti, D. (2003). Moving research on resilience into the 21st century: Theoretical and methodological considerations in examining the biological contributors to resilience. *Development and Psychopathology, 15*, 773–810.

Dahl, R., & Ryan, N. (1996). The psychobiology of adolescent depression. In D. Cicchetti & S. L. Toth (Eds.), *Rochester symposium on developmental psychopathology: Vol. 7. Adolescence: Opportunities and challenges* (pp. 197–232). Rochester, NY: University of Rochester Press.

DeBellis, M. D., Baum, A. S., Birmaher, B., Keshavan, M. S., Eccard, C. H., Boring, A. M., et al. (1999). Developmental traumatology Part I: Biological stress systems. *Biological Psychiatry, 45*, 1259–1270.

DeSantis, A. S., Adam, E. K., Doane, L. D., Mineka, S., Zinbarg, R. E., & Craske, M. G. (2007). Racial/ethnic differences in cortisol diurnal rhythms in a community sample of adolescents. *Journal of Adolescent Health, 41*, 3–13.

English, D. J., Upadhyaya, M. P., Litrownik, A. J., Marshall, J. M., Runyan, D. K., Graham, J. C., et al. (2005). Maltreatment's wake: The relationship of maltreatment dimensions to child outcomes. *Child Abuse & Neglect, 29*, 597–619.

Feder, A., Coplan, J. D., Goetz, R. R., Pine, D. S., Dahl, R. E., Ryan, N. D., et al. (2004). Twenty-four-hour cortisol secretion in prepubertal children with anxiety or depressive disorders. *Biological Psychiatry, 56*, 198–204.

Fenoglio, K. A., Brunson, K. L., & Baram, T. Z. (2006). Hippocampal neuroplasticity induced by early-life stress: Functional and molecular aspects. *Frontiers in Neuroendocrinology, 27*, 180–192.

Fisher, P. A., Stoolmiller, M., & Gunnar, M. R. (2007). Effects of a therapeutic intervention for foster preschoolers on daytime cortisol activity. *Psychoneuroendocrinology, 32*, 892–905.

Gibb, B. E., Wheeler, R., Alloy, L. B., & Abramson, L. Y. (2001). Emotional, physical, and sexual maltreatment in childhood versus adolescence and personality dysfunction in young adulthood. *Journal of Personality Disorders, 15*, 505–511.

Gunnar, M. R. (2003). Integrating neuroscience and psychological approaches in the study of early experiences. *Annals of the New York Academy of Sciences, 1008*, 238–247.

Gunnar, M. R., & Fisher, P. A. (2006). Bringing basic research on early experience and stress neurobiology to bear on preventive interventions for neglected and maltreated children. *Development and Psychopathology, 18*, 651–677.

Gunnar, M. R., & Quevedo, K. (2007). The neurobiology of stress and development. *Annual Review of Psychology, 58*, 145–173.

Gunnar, M. R., & Vazquez, D. (2001). Low cortisol and a flattening of expected daytime rhythm: Potential indices of risk in human development. *Development and Psychopathology, 13*, 515–538.

Gunnar, M. R., & Vazquez, D. (2006). Stress, neurobiology and developmental psychopathology. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology: Vol. 2. Developmental neuroscience* (2nd ed., pp. 533–577). New York: Wiley.

Gunnar, M. R., Wewerka, S., Frenn, K., Long, J. D., & Griggs, C. (2009). Developmental changes in HPA activity over the transition to adolescence: Normative changes and associations with puberty. *Development and Psychopathology*, *21*, 69–85.

Hart, J., Gunnar, M., & Cicchetti, D. (1996). Altered neuroendocrine activity in maltreated children related to symptoms of depression. *Development and Psychopathology*, *8*, 201–214.

Heim, C., & Nemeroff, C. B. (2001). The role of childhood trauma in the neurobiology of mood and anxiety disorders: Preclinical and clinical studies. *Biological Psychiatry*, *49*, 1023–1039.

Heim, C., Newport, D. J., Bonsall, R., Miller, A. H., & Nemeroff, C. B. (2001). Altered pituitary-adrenal axis responses to provocative challenge tests in adult survivors of childhood abuse: The role of comorbid depression. *American Journal of Psychiatry*, *158*, 575–581.

Heim, C., Newport, D. J., Mletzko, T., Miller, A. H., & Nemeroff, C. B. (2008). The link between childhood trauma and depression: Insights from HPA axis studies in humans. *Psychoneuroendocrinology*, *33*, 693–710.

Heim, C., Owen, M. J., Plotsky, P. M., & Nemeroff, C. B. (1997). Persistent changes in corticotrophin-releasing factor systems due to early life stress: Relationship to the pathophysiology of major depression and post-traumatic stress disorder. *Psychopharmacology Bulletin*, *33*, 185–192.

Heim, C., Plotsky, P. M., & Nemeroff, C. B. (2004). Importance of studying the contributions of early adverse experience to neurobiological findings in depression. *Neuropsychopharmacology*, *29*, 641–648.

Holsboer, F. (2000). The corticosteroid receptor hypothesis of depression. *Neuropsychopharmacology*, *23*, 477–501.

Kaplow, J. B., Dodge, K. A., Amaya-Jackson, L., & Saxe, G. N. (2005). Pathways to PTSD, part II: Sexually abused children. *American Journal of Psychiatry*, *162*, 1305–1310.

Kaplow, J. B., & Widom, C. S. (2007). Age of onset of child maltreatment predicts long-term mental health outcomes. *Journal of Abnormal Psychology*, *116*, 176–187.

Kaufman, J. (1991). Depressive disorders in maltreated children. *Journal of the American Academy of Child and Adolescent Psychiatry*, *30*, 257–265.

Kaufman, J., Birmaher, B., Perel, J., Dahl, R. E., Moreci, P., Nelson, B., et al. (1997). The corticotropic-releasing hormone challenge in depressed abused, depressed nonabused, and normal control children. *Biological Psychiatry*, *42*, 669–679.

Kaufman, J., Cook, A., Amy, L., Jones, B., & Pittinksy, T. (1994). Problems defining resiliency: Illustrations from the study of maltreated children. *Development and Psychopathology*, 6, 215–229.

Kaufman, J., Yang, B., Douglas-Palumberi, H., Houshyar, S., Lipschitz, D., Krystal, J., et al. (2004). Social supports and serotonin transporter gene moderate depression in maltreated children. *Proceedings of the National Academy of Sciences of the USA*, *101*, 17316–17321.

Keiley, M. K., Howe, T. R., Dodge, K. A., Bates, J. E., & Pettit, G. S. (2001). The timing of child physical maltreatment: A cross-domain growth analysis of impact on adolescent externalizing and internalizing problems. *Development and Psychopathology*, *13*, 891–912.

Kim, J., & Cicchetti, D. (2006). Longitudinal trajectories of self-system and depressive symptoms among maltreated and nonmaltreated children. *Child Development*, *77*, 624–639.

Kovacs, M. (1982). *The Children's Depression Inventory: A self-rated depression scale for school-aged youngsters.* Unpublished manuscript, Pittsburgh, PA: University of Pittsburgh.

Kovacs, M. (2004). *Children's Depression Inventory (CDI).* Toronto, Canada: Multi-Health Systems.

Levine, S. (2005). Developmental determinants of sensitivity and resistance to stress. *Psychoneuroendocrinology*, *30*, 939–946.

Maestripieri, D. (1999). The biology of human parenting: Insights from nonhuman primates. *Neuroscience & Biobehavioral Reviews*, *23*, 411–422.

Makino, S., Gold, P. W., & Schulkin, J. (1994). Corticosterone effects on corticotropin-releasing hormone mRNA in the central nucleus of the amygdala and the parvocellular region of the paraventricular nucleus of the hypothalamus. *Brain Research*, *640*, 105–112.

Manly, J. T., Cicchetti, D., & Barnett, D. (1994). The impact of subtype, frequency, chronicity, and severity of child maltreatment on social competence and behavior problems. *Development and Psychopathology*, *6*, 121–143.

Manly, J. T., Kim, J. E., Rogosch, F. A., & Cicchetti, D. (2001). Dimensions of child maltreatment and children's adjustment: Contributions of developmental timing and subtype. *Development and Psychopathology*, *13*, 759–782.

McEwen, B. (2000). Allostasis and allostatic load: Implications for neuropsychopharmacology. *Neuropsychopharmacology*, *22*, 108–124.

McEwen, B. (2003). Mood disorders and allostatic load. *Biological Psychiatry*, *54*, 200–207.

McEwen, B., & Stellar, E. (1993). Stress and the individual mechanisms leading to disease. *Archives of Internal Medicine*, *153*, 2093–2101.

Meaney, M. J., & Szyf, M. (2005). Environmental programming of stress responses through DNA methylation: Life at the interface between a dynamic environment and a fixed genome. *Dialogues in Clinical Neuroscience*, *7*, 103–123.

Miller, G. E., Chen, E., & Zhou, E. S. (2007). If it goes up, must it come down? Chronic stress and the hypothalamic-pituitary-adrenocortical axis in humans. *Psychological Bulletin*, *133*, 25–45.

Pears, K., & Fisher, P. A. (2005). Developmental, cognitive, and neuropsychological functioning in preschool-aged foster children: Associations with prior maltreatment and placement history. *Journal of Developmental and Behavioral Pediatrics*, 26, 112–122.

Pollak, S. D., Cicchetti, D., & Klorman, R. (1998). Stress, memory, and emotion: Developmental considerations from the study of child maltreatment. *Development and Psychopathology*, *10*, 811–828.

Pollak, S. D., & Sinha, P. (2002). Effects of early experience on children's recognition of facial displays of emotion. *Developmental Psychology*, *38*, 784–791.

Raison, C. L., & Miller, A. H. (2003). When not enough is too much: The role of insufficient glucocorticoid signaling in the pathophysiology of stress-related disorders. *American Journal of Psychiatry*, *160*, 1554–1565.

Rieder, C., & Cicchetti, D. (1989). Organizational perspective on cognitive control functioning and cognitive-affective balance in maltreated children. *Developmental Psychology*, *25*, 382–393.

Sanchez, M. M., Ladd, C. O., & Plotsky, P. M. (2001). Early adverse experience as a developmental risk factor for later psychopathology: Evidence from rodent and primate models. *Development and Psychopathology*, *13*, 419–450.

Sanchez, M. M., McCormack, K. M., & Maestripieri, D. (in press). Ethological case study: Infant abuse in rhesus macaques. In C. M. Worthman, P. M. Plotsky, D. S. Schechter, & C. Cummings (Eds.), *Formative experiences: The interaction of caregiving, culture, and developmental psychobiology*. New York: Cambridge University Press.

Sanchez, M. M., Noble, P. M., Lyon, C. K., Plotsky, P. M., Davis, M., Nemeroff, C. B., et al. (2005). Alterations in diurnal cortisol rhythm and acoustic startle response in nonhuman primates with adverse rearing. *Biological Psychiatry*, *57*, 373–381.

Schwartz, E. B., Granger, D. A., Susman, E. J., Gunnar, M. R. & Laird, B. (1999). Assessing salivary cortisol in studies of child development. *Child Development*, *69*, 1503–1513.

Sedlack, A. J., & Broadhurst, D. D. (1996). *The Third National Incidence Study of Child Abuse and Neglect (NIS-3) final report*. Washington, DC: U.S. Department of Health and Human Services.

Shackman, J. E., Shackman, A. J., & Pollak, S. D. (2007). Physical abuse amplifies attention to threat and increases anxiety in children. *Emotion*, *7*, 838–852.

Stroud, L., Foster, E., Handwerger, K., Papandonatos, G. D., Granger, D., Kivlighan, K. T., et al. (2009). Stress response and the adolescent transition: Performance versus peer rejection stress. *Development and Psychopathology*, *21*, 47–68.

Susman, E. J. (2006). Psychobiology of persistent antisocial behavior: Stress, early vulnerabilities and the attenuation hypothesis. *Neuroscience and Biobehavioral Reviews*, *30*, 376–389.

Susman, E. J., Dockray, S., Dorn, L. D., Schiefelbein, V. L., Herwehe, S., & Heaton, J. A. (2007). Morningness/eveningness, morning-to-afternoon cortisol ratio, and antisocial behavior problems during puberty. *Developmental Psychology*, *43*, 811–822.

Tarullo, A. R., & Gunnar, M. R. (2006). Child maltreatment and the developing HPA axis. *Hormones and Behavior*, *50*, 632–639.

Teisl, M., & Cicchetti, D. (2008). Physical abuse, cognitive and emotional processes, and aggressive/disruptive behavior problems. *Social Development*, *16*, 1–23.

Thompson, R. A., & Nelson, C. A. (2001). Developmental science and the media: Early brain development. *American Psychologist*, *56*, 5–15.

Toth, S. L., & Cicchetti, D. (1993). Child maltreatment: Where do we go from here in our treatment of victims? In D. Cicchetti & S. L. Toth (Eds.), *Child abuse, child development, and social policy* (pp. 399–438). Norwood, NJ: Ablex.

Toth, S. L., Manly, J. T., & Cicchetti, D. (1992). Child maltreatment and vulnerability to depression. *Development and Psychopathology*, *4*, 97–112.

U.S. Department of Health and Human Services, Administration on Children, Youth, and Families. (2000). *Child maltreatment 1998: Reports from the states to the National Child Abuse and Neglect Data System*. Washington, DC: U.S. Government Printing Office.

Widom, C. S., DuMont, K., & Czaja, S. J. (2007). A prospective investigation of major depressive disorder and comorbidity in abused and neglected children grown up. *Archives of General Psychiatry*, *64*, 49–56.

Zhang, L. X., Levine, S., Dent, G., Zhan, Y., Xing, G., Okimoto, D., et al. (2002). Maternal deprivation increases cell death in the infant rat brain. *Brain Research: Developmental Brain Research*, *133*, 1–111.

Child Development, January/February 2010, Volume 81, Number 1, Pages 270–289

Biological Sensitivity to Context: The Interactive Effects of Stress Reactivity and Family Adversity on Socioemotional Behavior and School Readiness

Jelena Obradović
Stanford University

Nicole R. Bush
University of California, San Francisco

Juliet Stamperdahl
University of California, Berkeley

Nancy E. Adler
University of California, San Francisco

W. Thomas Boyce
University of British Columbia, Vancouver

This study examined the direct and interactive effects of stress reactivity and family adversity on socioemotional and cognitive development in three hundred and thirty-eight 5- to 6-year-old children. Neurobiological stress reactivity was measured as respiratory sinus arrhythmia and salivary cortisol responses to social, cognitive, sensory, and emotional challenges. Adaptation was assessed using child, parent, and teacher reports of externalizing symptoms, prosocial behaviors, school engagement, and academic competence. Results revealed significant interactions between reactivity and adversity. High stress reactivity was associated with more maladaptive outcomes in the context of high adversity but with better adaption in the context of low adversity. The findings corroborate a reconceptualization of stress reactivity as *biological sensitivity to context* by showing that high reactivity can both hinder and promote adaptive functioning.

A substantive body of work has established that environmental adversity can have a deleterious effect on children's functioning (Luthar, 2006; Obradović, Shaffer, & Masten, in press; Rutter, 1983; Sameroff, 2006). Exposure to adverse, stressful events, such as marital conflict, maternal depression, and financial stress, has been linked to socioemotional behavior problems and cognitive deficits (Boyce, 2007; Boyce et al., 2001; Burchinal, Roberts, Hooper, & Zeisel, 2000; Cummings & Davies, 2002; Essex, Klein, Cho, & Kalin, 2002; Masten & Shaffer, 2006). Nevertheless, not all children are equally susceptible to adverse environmental influences. In accordance with the "stress diathesis" hypothesis, behaviorally or biologically reactive children have been traditionally identified as particularly vulnerable to stressful experiences, showing higher levels of behavioral and health problems than their low-reactive peers in the context of environmental risk and adversity (Belsky, Hsieh, & Crnic, 1998; Cummings, El-Sheikh, Kouros, & Keller, 2007; Deater-Deckard & Dodge, 1997; El-Sheikh, 2005; Ramos, Guerin, Gottfried, Bathurst, & Oliver, 2005). In recent years, however, a few researchers have challenged this traditional view, arguing that high reactivity, whether measured at the emotional, behavioral, or biological level, is not a unitary, pathogenic response to adversity that invariably leads to maladaptation. Belsky and colleagues (Belsky, Bakermans-Kranenburg, & van IJzendoorn, 2007), for example, have posited that temperamentally

---

This study was supported by a grant (R01 MH62320) from the National Institute of Mental Health. Preparation of this manuscript was in part supported by a Killam Postdoctoral Research Fellowship from the University of British Columbia to the first author. Dr. Obradović is the Great-West Life Junior Fellow in the Canadian Institute for Advanced Research's Experience-Based Brain and Biological Development Program and Junior Fellow Academy. Dr. Boyce is a Fellow of the Canadian Institute for Advanced Research and holds the Sunny Hill Health Centre/BC Leadership Chair in Child Development at the University of British Columbia. The authors acknowledge the substantive contributions made by Dr. Abbey Alkon to the development of our stress reactivity paradigm and the analysis of reactivity data. We are also appreciative of the school principals, teachers, children, and families who participated in the reported study and the numerous research assistants who collected and scored these data.

Correspondence concerning this article should be addressed to Jelena Obradović, Stanford University, School of Education, 485 Lasuen Mall, Stanford, CA 94305-3096. Electronic mail may be sent to jelena.obradovic@stanford.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8004-0018

reactive children may show higher susceptibility to environmental influences "for better and for worse." That is, reactive children may be more vulnerable to contextual risk, but they may also show more adaptive responses to interventions (Blair, 2002; Klein Velderman, Bakermans-Kranenburg, Juffer, & van IJzendoorn, 2006). Similarly, Boyce and colleagues have proposed a new theory suggesting that stress reactivity is better conceptualized as a high biological sensitivity to context (Boyce, 2007; Boyce & Ellis, 2005). From this theoretical perspective, children with heightened biological sensitivity to context are predicted to be more vulnerable to negative and stressful contextual factors, but also to have greater capacity to benefit from positive environmental influences. Thus, high biological sensitivity may be maladaptive in the context of adversity but adaptive in the context of a nurturing and supportive environment (Boyce, 1996; Boyce et al., 1995; Boyce et al., 2006; Ellis, Essex, & Boyce, 2005). This work underlines the importance of understanding more fully the biological processes that interact with environmental influences to shape children's adaptation, as indexed by competence and psychopathology (Curtis & Cicchetti, 2003; Masten & Obradović, 2006). The current study was designed to examine how early exposure to family adversity interacts with children's stress reactivity to predict indices of socioemotional behavior, school engagement, and academic competence.

Human Stress Response

Every day, in all human societies, children are exposed to challenging or stressful events from multiple sources. These events range from normative, potentially positive events, such as forming a new peer group; to adverse events, such as witnessing marital conflict; and truly traumatic events that directly threaten children's well-being, such as abuse and neglect. Children respond to these stressors with a set of highly integrated, neurobiological stress responses. A "fight-or-flight" response activates the autonomic nervous system (ANS), which initiates, within seconds, an integrated, short-onset repertoire of biobehavioral changes associated with accelerations of heart and respiratory rates, sweat production, and other physiological changes. The ANS comprises both a sympathetic (SNS) branch, which initiates physiological arousal, and a parasympathetic (PNS) branch, which modulates SNS input to the heart and other target organs, regulating recovery and restoring autonomic homeostasis (Berntson, Cacioppo, & Quigley, 1993). In a lagged response occurring within minutes, activation of the hypothalamic–pituitary–adrenocortical (HPA) axis via glucocorticoid secretion regulates glucose metabolism, blood pressure, and immunity and counterbalances the effects of the integrated ANS stress response (Sapolsky, Romero, & Munck, 2000). There are numerous ways of assessing children's stress reactivity. Because of our interest in whether stress reactivity moderates the effects of adversity on indices of psychopathology and competence, we examined two indicators of stress reactivity that are known to play a role in regulating stress-induced arousal and maintaining homeostasis: (a) respiratory sinus arrhythmia (RSA) and (b) salivary cortisol.

Respiratory Sinus Arrhythmia

RSA is a measure of parasympathetic stress response and refers to high-frequency heart rate variation controlled by efferent fibers of the vagus nerve during the respiratory cycle. Vagal regulation, in the form of increases and decreases in RSA, has been regarded as an index of children's capacity to regulate responses to positive and negative environmental demands (Beauchaine, 2001; Beauchaine, Gatzke-Kopp, & Mead, 2007; Porges, 2001, 2003, 2007). High levels of basal RSA have been associated with social competence, empathy, and emotion regulation (Beauchaine, 2001; Eisenberg et al., 1995; Fabes, Eisenberg, & Eisenbud, 1993; Fox & Field, 1989). Low basal RSA, on the other hand, may indicate emotional lability and dysregulation and has been linked to behavior problems in at-risk or clinical samples (Beauchaine, 2001; Beauchaine et al., 2007; Field et al., 1996; Mezzacappa et al., 1997; Pine et al., 1996; Pine et al., 1998). However, it is important to note that recent studies of community, nonclinical samples have found a positive relation, or no significant relation, between resting RSA and externalizing symptoms (Calkins, Graziano, & Keane, 2007; Dietrich et al., 2007).

Further, researchers have examined changes in RSA in response to stressful laboratory challenges. High reactivity, as indexed by decreases in RSA from basal levels (i.e., greater vagal withdrawal), has been associated with more sustained attention, better emotion regulation, and increased engagement during challenging tasks (Calkins, 1997; Suess, Porges, & Plude, 1994). However, in community samples of kindergarten children, higher RSA reactivity to various challenge tasks has been linked to seemingly contradictory indices of adaptation, such as high levels of sociability (Doussard-Roosevelt, Montogomery, & Porges, 2003) and internalizing symptoms (Boyce et al., 2001). Some of the inconsistencies across studies might be explained by the nature of the sample. For example, low RSA reactivity has been associated with externalizing symptoms in normative samples of young children (Boyce et al., 2001; Calkins et al., 2007), whereas high RSA reactivity in response to emotional stimuli has been observed in children with clinical levels of behavior problems (Crowell et al., 2005).

Only a few studies have examined the effect of RSA in the context of both high and low adversity. Several studies of community samples indicate that high basal RSA levels and high RSA reactivity in response to emotion-evoking stimuli (e.g., an adult argument) buffered children from the effects of marital conflict on behavior problems, academic achievement, and health (El-Sheikh, Harger, & Whitson, 2001; El-Sheikh & Whitson, 2006; Katz & Gottman, 1995, 1997). Only children with low baseline RSA and RSA reactivity were at risk for maladaptation in the context of high marital conflict. However, these findings varied across gender, type of marital conflict (e.g., verbal, physical), and children's outcomes. Similarly, high RSA reactivity to interpersonal stress emerged as a protective factor against the effects of hostile-withdrawn parenting on peer conflict but not on positive indices of peer interactions (Leary & Katz, 2004). In contrast, in a clinical sample, parental psychopathology had a negative effect on children's emotional and behavioral problems only in children with high baseline RSA (Shannon, Beauchaine, Brenner, Neuhaus, & Gatzke-Kopp, 2007). Studies examining other indices of ANS stress reactivity (i.e., heart rate, blood pressure, skin conductance) in community samples also indicate that children with high stress reactivity may be at risk for social, cognitive, behavioral, and health problems in the context of high adversity but may show better developmental outcomes in the context of low adversity (Boyce, 1996; Boyce et al., 1995; Boyce et al., 2006; Cummings et al., 2007; El-Sheikh, 2005).

In sum, the effect of stress reactivity on adaptation seems to be context dependent, but the directions of reactivity effects across different contexts remain unclear. Further studies are needed to elucidate the role of RSA reactivity in the development of competence and psychopathology, bearing in mind the limitations of previous research. First, many existing studies of RSA reactivity have important methodological flaws, such as a lack of appropriate controls for respiratory rate, attention, and motor activity (Beauchaine, 2001). Second, RSA reactivity has been examined mostly in relation to a single challenge, potentially misrepresenting the comprehensive reactivity to various types of challenges that children may encounter in their daily lives. Third, studies have often focused on a single source of adversity, such as marital conflict, yet children face various kinds of adversities in family environments. Fourth, studies of stress reactivity in general have been largely focused on measures of behavioral and health problems to the exclusion of more positive and prosocial indicators of adaptation.

Cortisol and Adaptive Functioning

Cortisol is the human glucocorticoid hormone secreted by the adrenal glands and passively diffuses into saliva. Salivary cortisol reflects the levels of unbound and biologically active cortisol circulating in the blood. As with RSA, researchers have linked individual differences in both basal and reactive cortisol expression primarily to indices of behavior problems (Gunnar & Vazquez, 2006). In general, in both clinical and community samples, elevated daily cortisol levels have been associated with internalizing symptoms (Goodyer, Herbert, & Althan, 1998; Klimes-Dougan, Hastings, Granger, Usher, & Zahn-Waxler, 2001), whereas lower levels of daily cortisol have been associated with externalizing symptoms (Hardie, Moss, Vanyukov, Yao, & Kirillovac, 2002; King, Barkley, & Barrett, 1998; McBurnett, Lahey, Rathouz, & Loeber, 2000; McBurnett et al., 1991; Moss, Vanyukov, & Martin, 1995; Shirtcliff, Granger, Booth, & Johnson, 2005). However, in community samples of kindergarten children, basal cortisol secretion was positively related to internalizing symptoms, social wariness, and symptom severity (Essex et al., 2002; Smider et al., 2002).

A few studies have examined cortisol reactivity to specific laboratory challenges. Moderate elevation of cortisol in response to cognitive tasks has been associated with better executive functioning and self-regulation skills in both middle- and low-income samples (Blair, Granger, & Razza, 2005; Davis, Bruce, & Gunnar, 2002). In addition, high cortisol reactivity in response to parent–child conflict tasks has been associated with social problems and internalizing symptoms in clinical samples (Granger, Weisz, & Kauneckis, 1994; Granger, Weisz, McCracken, Ikeda, & Douglas, 1996). However, differences in sample characteristics and stimuli across these studies preclude any conclusions regarding the relation between cortisol reactivity to different challenges and adaptive functioning.

Studies examining cortisol changes over a day in child care or preschool in community samples have linked elevated cortisol to impulsivity, poor effortful control, peer rejection, and aggression (Dettling, Gunnar, & Donzella, 1999; Dettling, Parker, Lane, Sebanc, & Gunnar, 2000; Gunnar, Sebanc, Tout, Donzella, & van Dulmen, 2003). Similarly, elevated cortisol in response to starting a new school year has been associated with negative affectivity, and surgency and extraversion (Davis, Donzella, Krueger, & Gunnar, 1999). However, a more complex picture emerged from a study that measured children's cortisol response during the first weeks of school, when children forge new peer groups and connections, and again later in the school year, when peer groups have been fully formed (Gunnar, Tout, de Haan, Pierce, & Stansbury, 1997). Children showing elevated cortisol expression during the first weeks of school and normal levels later in the year were rated as extraverted, socially competent, and outgoing, whereas children who showed high levels of cortisol later in the school year showed more solitary and negative behaviors and were seen as less competent and less outgoing.

Although these studies indicate that the relation between cortisol reactivity and adaptation may be context dependent and may vary with levels of contextual stress, there is a paucity of studies examining the interactive effects of cortisol and adversity on adaptation. We are aware of only one such study, which shows that cortisol reactivity is related to symptom severity only for children with low father involvement in infancy (Boyce et al., 2006). Moreover, studies examining cortisol response to various age-appropriate stressors and in the context of overall family adversity are rare. Thus, it is unknown whether cortisol reactivity may diminish or amplify the effects of adversity on children's adaptation. In order to better understand the regulatory function of cortisol, researchers need to examine whether cortisol reactivity interacts with contextual factors in predicting both positive and negative developmental outcomes.

Current Study

The main goal of this study was to investigate how the interplay between children's stress reactivity and overall family adversity influences adaptation. We focused our study on kindergarten children because understanding processes that subserve adaptation during the salient developmental transition to a school setting is very important, given the reverberatory effects that early behavior problems or academic failure can have on later developmental trajectories. Consequently, we examined adaptation across four domains of functioning known to contribute to early school success and later adjustment: (a) externalizing symptoms, (b) prosocial behaviors toward peers, (c) school engagement, and (d) academic competence.

Based on the broad literature on risk and adversity, we hypothesized a robust negative main effect of family adversity across all indices of adaptation. We also expected to find main effects of stress reactivity on adaptation, but given the simultaneous test of interactive effects, as well as the paucity of studies examining the effects of RSA and cortisol reactivity, especially for positive developmental outcomes, and some inconsistencies within such studies, we did not hypothesize the directions of these main effects. More importantly, in accordance with the theory of biological sensitivity to context (Boyce & Ellis, 2005; Ellis et al., 2005), we expected to find evidence that ANS and HPA reactivity moderate the effects of early family adversity on various domains of functioning. We hypothesized that in high-adversity family environments, elevated levels of stress reactivity would be associated with maladaptive outcomes, whereas low stress reactivity would act as a protective factor. In the context of low family adversity, on the other hand, we expected high levels of reactivity to be associated with better adaptation. It is important to note that although biological sensitivity to context should be examined in both positive and negative settings, our assessment focuses on six types of family adversities, and a lack of overall family adversity does not necessarily imply a supportive and nurturing environment. In addition to the hypothesized Adversity × Stress Reactivity interactions, we controlled for children's sex and tested whether main and interactive effects of adversity and reactivity vary across sex.

Method

Participants

The sample included 338 children (163 females, 175 males) who participated in a longitudinal study of social status, biological responses to adversity, and child mental and physical health. Participants were recruited in three waves from 29 kindergarten classrooms within six public schools in the San Francisco Bay Area (Oakland, Albany, and

Piedmont Unified school districts) during the falls of 2003, 2004, and 2005. Schools were selected to represent a variety of sociodemographic and ethnic characteristics of the metropolitan area. Families were recruited through a series of efforts that included home mailings, presentations at kindergarten parent welcome nights, and in-person recruitment during drop-off and pick-up. Every family with a child in the target classrooms was invited to participate. However, families who were not fluent in either English or Spanish were excluded to ensure adequate comprehension of study questionnaires. Schools were provided with $20 per child enrolled in the study.

The mean age of the children at the kindergarten entry was 5.32 years (*SD* = 0.32, range = 4.75–6.28). The sample was ethnically diverse, and children were identified as 19% African American, 11% Asian, 43% European or White, 4% Latino, 22% multiethnic, and 2% as "other" ethnicity. Family demographic information was not provided by 16 families. Primary caregivers were 87% biological mothers, 9% biological fathers, 2.5% adoptive mothers, 0.6% biological grandmothers, and 0.9% "other" relations. As the vast majority of primary caregivers were parents, these terms are used interchangeably here. Average annual household income ranged from < $10,000 to > $400,000 (*M* = $60,000–79,999; *Mdn* = $80,000–99,999). Highest level of educational attainment in the household ranged from less than a high school diploma to advanced degrees, with 75% of caregivers having at least a college degree.

Procedures

Data for this study were collected in the fall (Time 1) and spring (Time 2) during the kindergarten year. Parents' informed consent and children's assent were obtained prior to the start of data collection. Parent report of family adversity and children's functioning was collected through a series of home mailings, and families were compensated with $50 for each completed time point. Teacher report of children's functioning was collected through questionnaires dropped off and picked up at the child's school, and teachers were compensated $15 per child for each report returned. Self-report of children's functioning was collected through a series of structured, scripted interviews conducted in a separate, quiet room at elementary schools.

In addition, children participated in the 20-min reactivity protocol, which included four age-appropriate tasks designed to elicit stress responses to social, cognitive, sensory, and emotional challenges (Alkon et al., 2003; Boyce et al., 1995). Since PNS can be activated by psychomotor activity, such as gesturing, speaking, focused attending, and nonchallenging social engagement (Bazhenova & Porges, 1997; Bernardi et al., 2000; Berntson, Cacioppo, & Fieldstone, 1996; Porges, 2001; Porges et al., 2007), RSA assessed during a challenge task reflects both stress reactivity in response to the challenge and reactivity in response to the psychomotor or engagement demands of the challenge task. The recent literature has emphasized the importance of maximizing the mobilization of the psychological response to the particular challenge and minimizing the peripheral triggers of cardiovascular activation (Bush, Alkon, Obradović, Stamperdahl, & Boyce, 2009; Kamarck & Lovallo, 2003). Thus, in order to best measure the psychophysiological response to the four challenges, each challenge, task in the reactivity protocol was preceded by a nonchallenging "control task" that paralleled the motor and engagement demands of that challenge, and levels of autonomic arousal during the control tasks were used as baseline values.

The autonomic reactivity protocol started with an experimenter reading a calming short story (2 min). Next, the child completed four sets of paired tasks, each consisting of a control condition and challenge condition. The *social challenge task* (2 min) was a structured interview about the child's family, friends, and likes/dislikes adapted from the Gesell School Readiness Screening Test (Carlson, 1985). This was preceded by the *social control task* in which children were asked to name common animals and colors from a picture book to capture arousal associated with speaking, gestures during social speech, and attention involved in social engagement (2 min). The *cognitive challenge task* (2 min) was a digit span recitation task derived from the Kaufman Assessment Battery for Children (Kaufman & Kaufman, 1983) in which children were asked to recall sequences of numbers up to six digits in length and received negative verbal feedback after making a mistake. This was preceded by the *cognitive control task* in which children were asked to repeat simple, one- or two-digit number sequences to capture arousal associated with listening, speaking numbers, and social engagement (1 min). The *sensory challenge task* (1 min) was a taste-identification task (Kagan & Snidman, 1991) during which children were asked to identify two drops of concentrated lemon juice placed on their tongues. This was preceded by the *sensory control task* in which children were asked to identify two drops of water placed on their tongues to capture arousal associated with mouth opening and swallowing, anticipation, and guessing the content of the liquid (1 min). The *emotional challenge task* (2 min) consisted of watching a short emotion-evoking movie clip chosen to elicit fear (Eisenberg et al., 1988). This was preceded by the *emotion-control task* in which children were asked to watch an emotionally neutral movie clip to capture physiological responding associated with attending to visual stimuli (2 min). The autonomic reactivity protocol concluded with the reading of another calming story (2 min).

Measures

Stress reactivity. Children's stress reactivity was assessed using changes in both RSA and *salivary cortisol* during the stress reactivity protocol. After the child was familiarized with the equipment, four spot electrodes (two current, two impedance) were placed in the standard tetrapolar configuration on the child's neck and chest, and ECG electrodes were placed on the right clavicle and lower left rib. A 4 mA AC current at 100 kHz was passed through the two current electrodes. RSA was monitored continuously during the protocol. Data were acquired using the Biopac MP150 (Biopac Systems, Santa Barbara, CA) interfaced to a PC-based computer. Analog data were continuously monitored on the computer for signal and noise, and digitized data were stored for offline analysis. RSA was derived in accordance with recommendations of the Society for Psychophysiological Research committee on heart rate variability (Berntson et al., 1997). The sampling frequency was 1 kHz. Prior to analyses, each waveform was verified, IBIs were visually checked, and artifacts were identified using Berntson, Quigley, Jang, and Boysen's (1990) algorithm within the MindWare software program (http://www.mindwaretech.com). RSA was estimated as the natural logarithm of the variance of heart period within the high-frequency bandpass associated with respiration at this age (i.e., 0.15–0.80 Hz; Bar-Haim, Marshall, & Fox, 2000; Rudolph, Rudolph, Hostetter, Lister, & Siegel, 2003). Outlier data were checked and verified minute by minute if they were > 3 *SD* from the group mean. Mean RSA magnitude was calculated for each 1-min interval and averaged within each task (Cacioppo, Uchino, & Berntson, 1994).

RSA reactivity responses during each of the control tasks were used as baseline reference values to create four task RSA reactivity scores. To control for the influence of baseline levels of arousal, we created four standardized residual scores as measures of reactivity to social, cognitive, sensory, and emotional challenges by regressing RSA values during the challenge tasks on the respective control task RSA values. Residuals reflect the extent to which an individual's physiologic response deviated from the regression line derived from sample responses (Manuck, Kasprowicz, & Muldoon, 1990). The four standardized residual scores were then averaged to create one index of RSA stress reactivity. Elevated RSA reactivity was represented by negative residual scores, indicating a decrease in RSA and vagal withdrawal in response to a given challenge. Conversely, low reactivity was represented by positive residual scores, indicating an increase in RSA and vagal activation in response to the challenge.

Descriptive statistics for RSA responses to tasks are presented in Table 1. Average RSA differences between challenge and control tasks were −0.25 (*SD* = 0.66) for the social challenge, 0.60 (*SD* = 0.67) for the cognitive challenge, −0.32 (*SD* = 0.84) for the sensory challenge, and 0.10 (*SD* = 0.47) for the emotional challenge task. RSA reactivity levels were significantly different from zero during the social ($t = -6.99$, $p < .001$), cognitive ($t = 16.40$, $p < .001$), sensory ($t = -7.05$, $p < .001$), and

Table 1
Descriptive Statistics for Stress Reactivity and Adversity Variables

Variable	Scale	Min.	Max.	*M*	*SD*
RSA					
Social control		3.29	9.65	6.81	1.12
Social challenge		3.47	10.26	6.56	1.10
Cognitive control		2.62	9.69	6.12	1.20
Cognitive challenge		3.35	9.82	6.72	1.10
Sensory control		2.77	9.91	6.98	1.13
Sensory challenge		3.04	10.90	6.66	1.20
Emotional control		2.74	10.15	7.03	1.12
Emotional challenge		2.48	10.17	7.13	1.10
Cortisol					
Preprotocol		0.10	24.40	5.10	3.51
Postprotocol		0.10	30.50	4.70	3.93
Adversity					
Financial stress	1–5	1.00	5.00	2.42	0.93
Parenting overload	1–5	1.20	5.00	3.12	0.68
Marital conflict	1–5	1.00	3.10	1.74	0.38
Family expressiveness	1–9	1.20	7.20	4.03	0.99
Anger expression	0–10	0.25	6.13	2.42	0.84
Maternal depression	1–4	1.00	3.00	1.37	0.30
Harsh/restrictive parenting	1–7	1.76	6.56	3.66	0.75

emotional ($t = 4.04$, $p < .001$) tasks. Averaged across the four challenge tasks, 47.1% of the sample showed a decrease in RSA from the levels elicited by the control tasks, suggesting that about half of the sample showed vagal withdrawal in response to challenges. Thus, the reactivity protocol effectively captured variability in child RSA responses to challenges.

At the beginning and end of the reactivity protocol, salivary cortisol was collected by instructing the child to chew on a cotton roll for 20–30 s. Wet cotton rolls were deposited in salivette tubes and stored at −7 °C until assayed. The cortisol samples were assayed using a commercial immunoassay with chemiluminescence detection (Cortisol Luminescence Immunoassay; IBL-Hamburg, Hamburg, Germany). The detection limit of the assay was 0.41 nmol/L. The mean interassay and intra-assay variations were 8.5% and 6.1%, respectively. Cortisol values above 55 nmol/L (< 1% of samples) were considered unreliable data and were discarded. Ten children in this sample were taking medications, such as human growth hormone and exogenous glucocorticoids, known to alter salivary cortisol levels (Masharani et al., 2005). These children were excluded from analyses of cortisol reactivity. Given that cortisol levels in saliva reach their peak approximately 15–20 min following stressor onset, cortisol values collected at the beginning of the session were considered baseline reference values, as they indexed children's cortisol expression in the familiar context of the kindergarten classroom. The average session lasted 27 min (SD = 3 min, range = 19–38 min). Thus, cortisol values collected at the end of the session were considered a measure of reactivity to a novel, mildly stressful context (e.g., strange experimenter, electrodes, challenge tasks), at the midpoint of the reactivity protocol. Descriptive statistics for pre- and postprotocol values are presented in Table 1. We created standardized residual scores by regressing postprotocol cortisol values on preprotocol, baseline values. Because cortisol collection occurred both in the morning and in the afternoon, we examined whether cortisol reactivity varied across the day, due to the known circadian activity of the HPA system. The time of collection was related to absolute preprotocol values ($r = .18$, $p < .01$) but not to postprotocol values and cortisol reactivity as indexed by standardized residual scores, which is consistent with the current literature (Kudielka, Schommer, Hellhammer, & Kirschbaum, 2004). Length of the session also did not significantly relate to cortisol reactivity. Mean cortisol levels pre- and postprotocol were 5.10 nmol/L ($SD = 3.51$) and 4.70 nmol/L ($SD = 3.93$), respectively. Mean difference between pre- and postprotocol values was −0.40 nmol/L ($SD = 3.81$). Positive cortisol reactivity values indicate a stress response, and 37.5% of the sample showed such a response to the challenge protocol. The average difference between pre- and postprotocol values for the responder group was 2.58 nmol/L ($SD = 3.86$).

Adversity. Children's exposure to adversity was assessed by six indices of potential sources of family stress. All measures of adversity were based on parent report, and reliability statistics are reported for the current sample. *Financial Stress* was assessed with four items derived from Essex et al. (2002) that measured parents' thoughts about money problems, difficulty paying bills, and limited opportunities due to lack of finances ($\alpha = .81$) *Parenting Overload* was assessed with five items derived from Essex et al. that measured feelings of being overwhelmed with parenting duties, juggling conflicting obligations, and lacking time to rest or pursue desired activities ($\alpha = .79$). *Marital Conflict* was assessed using the 10-item O'Leary–Porter Overt Hostility Scale ($\alpha = .72$) that measured how often parents openly argue, display physical and verbal hostility, and criticize each other in the presence of their children (Johnson & O'Leary, 1987; Porter & O'Leary, 1980). Exposure to *Negative/Anger Expressiveness* in the family was assessed using both the Family Expressiveness Questionnaire (FAQ; Halberstadt, 1986) and the Anger Expression Inventory (AEI; Spielberger, 1988). The FAQ consists of a 10-item negative dominant subscale ($\alpha = .83$), measuring the frequency of overt anger, contempt, and hostility among family members, and a 10-item negative subdominant subscale ($\alpha = .75$), measuring the frequency of passive sulking, crying, and disappointment among family members. The two FAQ subscales were averaged ($r = .55$, $p < .001$) to yield one measure of negative family expressiveness. The total AEI score was calculated using three 8-item subscales that assessed parents' tendency to express overtly towards other people ($\alpha = .69$), hold angry feelings inside ($\alpha = .68$), and control the experience and expression of anger ($\alpha = .74$). The overall scores based on FAQ and AEI were standardized and averaged ($r = .47$, $p < .001$) into one indicator of exposure to negative/anger expressiveness. *Maternal Depression* was assessed with a 20-item Center for Epidemiological Studies Depression Scale (CES–D; $\alpha = .81$; Radloff, 1977). *Harsh and Restrictive Parenting* was assessed using a questionnaire version of Child-Rearing Practice Report

(CRPR; Block, 1965). The 18-item scale ($\alpha = .83$) was based on previous studies (Deković, Janssens, & Gerris, 1991; Rickel & Biasatti, 1982). Descriptive statistics for adversity measures are shown in Table 1. As we were interested in capturing children's overall exposure to family adversity, the six indices of adversity were standardized and composited into one adversity index. Although a few adversity measures included positive items, which provided some assessment of positive family environment (e.g., financial security, parental affection, parent who is happy, has time to relax, shows patience, and controls anger), it is important to note that, low levels of adversity on this measure do not necessarily indicate a supportive family environment.

Adaptive functioning. Adaptive functioning was assessed during the fall and spring of the kindergarten year using parent-, teacher-, and child-reported measures. For the purposes of this study, we focused on four developmentally salient domains—externalizing symptoms, prosocial behaviors, school engagement, and academic competence—which are described in more detail next. Parent and teacher perceptions of the child's behaviors and skills across these domains were assessed using The MacArthur Health and Behavior Questionnaire (HBQ; Armstrong, Goldstein, & the MacArthur Working Group on Outcome Assessment, 2003). Parent and teacher forms of the HBQ contain parallel scales, described in detail next. Children's perceptions of their own behavior and skills across four domains were assessed using the Berkeley Puppet Interview, which is designed to elicit children's responses to statements that parallel the parent and teacher HBQ scales (BPI; Ablow, Measelle, & the MacArthur Working Group on Outcome Assessment, 2003). The BPI combines both structured and clinical interviewing techniques and allows young children to respond either verbally or nonverbally to opposing statements issued by two puppets about their behavior. Interviewers were extensively trained to probe for more information, to deal with unclear, inaudible, indecisive, or alternate responses, and to create a comfortable, engaging, and interactive environment. Videotaped responses were coded using a 7-point scale based on the statement the child chose to endorse and the degree of endorsement. Interclass correlation coefficients (ICC) were calculated based on 35% of double-coded fall BPIs (ICC range = .91–.98) and 29% of double-coded spring BPIs (ICC range = .92–.99), indicating excellent interrater reliability. Individual items were averaged into subscales and scales according to the BPI manual. The validity and reliability of HBQ and BPI measures have been tested with children ages 5–6, and all three measures show good measurement properties (for details see Ablow et al., 2003; Armstrong et al., 2003).

All three measures contained the same scales. *Externalizing Symptoms* were assessed using an Oppositional Defiant scale (e.g., argues, angry, resentful, misbehaves), Conduct Problem scale (e.g., lies, vandalizes, threatens, cheats), and an Overt Hostility scale (e.g., teases, fights, kicks, taunts). Preliminary analyses revealed that indices of stress reactivity were related differently to a fourth Relational Aggression scale in comparison to the other three scales of externalizing symptoms. We believe that this finding deserves further investigation outside the scope of this study; we thus omitted the Relational Aggression scale from our measure of Externalizing Symptoms. *Prosocial Behaviors* were assessed using a scale measuring the child's predisposition to help others, share, invite bystanders to play, and be considerate, collaborative, fair and empathic. *School Engagement* was assessed using a scale measuring the extent to which the child was excited, eager, happy, frustrated, bored, or irritable about school. *Academic Competence* was assessed with a scale indexing the child's achievement, innate ability, and difficulties in literacy and math, as well as the child's overall academic performance in the classroom in comparison to other classmates. The sample's descriptive and reliability statistics for scales and domains across three informants and two time periods are presented in Table 2.

Using HBQ and BPI data, we conducted eight principal component analyses (PCA) to obtain multi-informant indices of externalizing symptoms, prosocial behaviors, school engagement, and academic competence, separately for the fall and spring assessments. We used a procedure described by Kraemer et al. (2003) for integrating data from multiple informants by triangulating three potentially orthogonal sources of information (i.e., parent, teacher, child). Following this approach, we used parent, teacher, and child report to extract three components for each PCA that conceptually correspond to the following dimensions: (a) trait (i.e., individual differences in adaptation), (b) perspective (i.e., informant characteristics that influence the assessment of the trait), and (c) context (i.e., environmental features that influences expression of the trait). For the purposes of this study, we used only factor scores based on the "trait" component, which weighs each source of information in

Table 2
Adaptive Functioning Measures

			Fall				Spring			
Domains/scales	Scale	No. items	*M*	*SD*	α	λ	*M*	*SD*	α	λ
Externalizing behavior problems										
P–HBQ			0.29	0.21	.78	.71	0.30	0.23	.80	.75
Oppositional defiant	0–2	9	0.44	0.31	.77		0.46	0.33	.80	
Conduct problems	0–2	12	0.13	0.16	.70		0.14	0.18	.77	
Overt hostility	0–2	4	0.27	0.27	.56		0.28	0.30	.66	
T–HBQ			0.19	0.30	.89	.71	0.26	0.36	.91	.76
Oppositional defiant	0–2	9	0.21	0.33	.88		0.29	0.41	.92	
Conduct problems	0–2	11	0.10	0.22	.73		0.14	0.26	.88	
Overt hostility	0–2	4	0.25	0.41	.82		0.33	0.47	.85	
C–BPI			2.40	0.55	.81	.67	2.36	0.51	.79	.73
Oppositional defiant	1–7	6	2.46	0.64	.49		2.41	0.63	.53	
Conduct problems	1–7	9	2.39	0.60	.65		2.35	0.55	.65	
Overt hostility	1–7	7	2.32	0.65	.72		2.30	0.64	.69	
Prosocial behaviors										
P–HBQ	0–2	20	1.31	0.31	.86	.68	1.29	0.32	.88	.70
T–HBQ	0–2	20	1.22	0.48	.95	.64	1.25	0.51	.95	.80
C–BPI	1–7	7	5.34	0.83	.64	.74	5.42	0.79	.67	.59
School engagement										
P–HBQ	1–4	8	3.73	0.35	.83	.65	3.65	0.42	.87	.75
T–HBQ	0–2	8	1.80	0.28	.81	.68	1.78	0.30	.84	.65
C–BPI	1–7	8	5.24	1.03	.80	.76	5.27	0.93	.77	.71
Academic competence										
P–HBQ	1–7	8	5.01	1.03	.92	.85	5.19	1.03	.93	.89
T–HBQ	1–5	5	3.34	0.80	.95	.75	3.58	0.92	.97	.86
C–BPI	1–7	12	5.13	0.80	.71	.53	5.16	0.75	.73	.52

Note. λ = factor loading; C–BPI = child report on Berkeley Puppet Interview; P–HBQ = parent report on Health and Behavioral Questionnaire; T–HBQ = teacher report on Health and Behavior Questionnaire.

the same direction, as an overall core measure of each adaptive domain. Thus, each adaptive domain (e.g., externalizing) is measured by standardized factor scores that represent the common variance across three reports of the domain and are largely free of informant and contextual effects. Table 2 shows PCA loadings for the first (trait) component of each adaptation domain in fall and spring.

Data Preparation and Analytic Procedures

Missing data for the six adversity components, the BPI and HBQ scales, pre- and postprotocol cortisol values, and RSA values to the four challenge and four control tasks were handled using the recommended maximum-likelihood estimation procedure for missing data, specifically, the expectation-maximization (EM) algorithm (Schafer & Graham, 2002). Percentages of missing data were as follows: fall teacher report (0.0%), fall child report (0.9%–1.5%), fall parent report (13.3%–13.9%), spring teacher report (3.0%), spring child report (3.6%–4.4%), spring parent report (16.9%–17.5%), adversity components (11.8%–13.0%), RSA values (3.6%–4.7%), and cortisol values (7.7%–9.8%). Extreme values on standardized final composites were truncated to −3.50 or 3.50 standardized values. All analyses were conducted using the SPSS 15.0 program (SPSS, Chicago). All hierarchical multiple regression analyses were conducted controlling for the child's sex and age at kindergarten entry.

Results

Bivariate Correlations

Bivariate correlations among key variables included in this study are presented in Table 3. Age at kindergarten entry was not related to indices of stress reactivity, adversity, or adaptation, and thus was not included in further analyses. Boys in

Table 3
Bivariate Correlations Among Age, Gender, Stress Reactivity, and Adaptive Functioning

	1	2	3	4	5	6	7	8	9	10	11	12
	Age	Sex	RSA	CORT	ADV	EXT–F	PRO–F	SE–F	ACA–F	EXT–S	PRO–S	SE–S
1 Age	—											
2 Sex	−.16**	—										
3 RSA	.02	.00	—									
4 CORT	−.06	.00	−.02	—								
5 ADV	.02	−.15**	.05	−.02	—							
6 EXT–F	.05	−.27***	.12*	.14*	.29***	—						
7 PRO–F	.00	.35***	.06	−.03	−.09	−.50***	—					
8 SE–F	.00	.21***	−.04	−.13*	−.25***	−.52***	.45***	—				
9 ACA–F	.00	.14**	−.07	−.09	−.02	−.17**	.38***	.35***	—			
19 EXT–S	.07	−.31***	.08	.10	.31***	.75***	−.46***	−.46***	−.14**	—		
11 PRO–S	−.01	.39***	.07	−.06	−.19***	−.48***	.71***	.43***	.24***	−.61***	—	
12 SE–S	.03	.19***	.00	−.14*	−.26***	−.47***	.44***	.59***	.22***	−.55***	.54***	—
13 ACA–S	.04	.11*	−.05	−.09	.01	.13*	.31***	.28***	.74***	−.16**	.28***	.30***

Note. ACA = academic competence; CORT = Cortisol reactivity; EXT = externalizing symptoms; F = fall; PRO = prosocial behavior; RSA = respiratory sinus arrhythmia reactivity; S = spring; SE = school engagement.
*$p < .05$. **$p < .01$. ***$p < .001$.

this sample were older than girls, experienced higher levels of adversity, and showed poorer adaptation across all domains. RSA reactivity was significantly related only to fall externalizing, with low reactive children exhibiting higher levels of symptoms. Cortisol reactivity was significantly related to fall externalizing symptoms and to both fall and spring school engagement, with high-reactive children exhibiting higher levels of symptoms and lower levels of engagement. Adversity exposure was related to higher externalizing symptoms and lower school engagement across both time points, as well as to lower prosocial behaviors in spring. Lastly, all domains of adaptation showed significant within-time correlations with other domains, as well as significant longitudinal stability from fall to spring.

Regression Analyses

The main goal of the regression analyses was to examine the interactive effects of adversity and stress reactivity on indices of adaptive functioning. While we also tested the main effects of family adversity and stress reactivity on adaptation, these direct effects can only be interpreted in the context of the significant interactive effects that qualify them. For each measure of stress reactivity (i.e., RSA and cortisol), two sets of regression analyses were conducted to predict (a) concurrent fall adaptation and (b) longitudinal change in adaptation by predicting spring adaptation while controlling for fall adaptation. Interactive effects were tested following the procedure outlined by Baron and Kenny (1986). Significant interactions were further investigated using the simple slopes technique proposed by Aiken and West (1991) by comparing the effect of high (i.e., 1 *SD* above the mean) and low (i.e., 1 *SD* below the mean) adversity in children with high and low stress reactivity. High RSA reactivity (i.e., vagal withdrawal) was defined as 1 *SD* below the mean, while low RSA reactivity was defined as 1 *SD* above the mean. In contrast, high cortisol reactivity was defined as 1 *SD* above the mean, and low cortisol reactivity was defined as 1 *SD* below the mean. Finally, we examined whether these main and interactive effects varied for boys and girls.

RSA reactivity. Results of regression analyses examining the effects of RSA reactivity and adversity on adaptation are presented in Table 4. Compared to boys, girls showed better adaptive functioning across all four domains in fall, as well as a decrease in externalizing symptoms and increase in prosocial behaviors across the 1st year of kindergarten. Lower RSA reactivity, as indexed by positive RSA residual values, significantly predicted higher levels of externalizing in fall. Higher adversity exposure predicted higher levels of externalizing and lower school engagement in fall. In addition, adversity exposure predicted an increase in externalizing symptoms as well as a decrease in

Table 4
Standardized Regression Coefficients for Models Testing the Effects of RSA Reactivity and Adversity on Adaptation

	Externalizing		Prosocial behavior		School engagement		Academic competence	
	Fall	Spring	Fall	Spring	Fall	Spring	Fall	Spring
Fall adaptation		.69***		.64***		.53***		.73***
Sex	−.22***	−.11**	.33***	.15***	.17**	.06	.13*	.00
RSA	.11*	−.01	.08	.04	−.03	.02	−.04	.02
ADV	.25***	.10**	−.05	−.11**	−.23***	−.12**	−.02	.00
RSA × ADV	−.11*	−.01	.11*	.03	.12*	.05	.07	.13***
ADV × Sex	−.07	−.06	.01	.02	.09	.03	−.10	.01
RSA × Sex	.03	−.05	.02	.05	.02	.00	.10	.08*
RSA × ADV × Sex	.02	−.03	.03	.01	−.02	−.08	−.03	.01
Total R^2	.17***	.60***	.14***	.54***	.12***	.37***	.05*	.57***

Note. ACA = academic competence; ADV = adversity; EXT = externalizing behavior problems; PRO = prosocial behavior; RSA = respiratory sinus arrhythmia reactivity; SE = school engagement.
*$p < .05$. **$p < .01$. ***$p < .001$.

prosocial behavior and school engagement from fall to spring. However, these main effects were qualified by significant interactions.

The interaction effect between RSA reactivity and adversity significantly predicted fall adaptation across three of four domains: externalizing symptoms, prosocial behaviors, and school engagement. Examination of simple slopes revealed that the effect of adversity exposure on adaptation varied across different levels of RSA reactivity. A stronger relation between adversity and externalizing symptoms was found for children with high RSA reactivity ($b = .35$, $p < .001$) than for those with low RSA reactivity ($b = .15$, $p < .05$). In the context of high adversity, both low- and high-reactivity groups showed elevated externalizing symptom levels, whereas in the context of low adversity exposure, high-reactive children had lower levels of symptoms than low reactive children (see Figure 1). Similarly, adversity exposure was significantly related to prosocial behavior in high-reactive children ($b = -.16$, $p < .05$) but not in low-reactive children ($b = .05$, $p > .05$). However, contrary to findings for externalizing symptoms, high-reactive children showed lower levels of prosocial behaviors than low-reactive children in the context of high adversity but did not differ in the context of low adversity (see Figure 2). Finally, adversity exposure was significantly related to fall school engagement in high-reactive children ($b = -.33$, $p < .001$) but not in low-reactive children ($b = -.12$, $p > .05$). In comparison to low-reactive peers, high-reactive children showed lower levels of fall school engagement in the context of high adversity but higher levels of school engagement in the context of low adversity (see Figure 3). High reactivity thus operated as a risk factor under conditions of high adversity but as a promotive, engagement-enhancing factor in circumstances of low adversity.

Although the interaction between RSA reactivity and adversity exposure did not predict fall levels of academic competence, it did predict change in academic competence from fall to spring. Analyses of simple slopes revealed that exposure to adversity was related to an increase in academic competence in low-reactive children ($b = .12$, $p < .05$) but to a

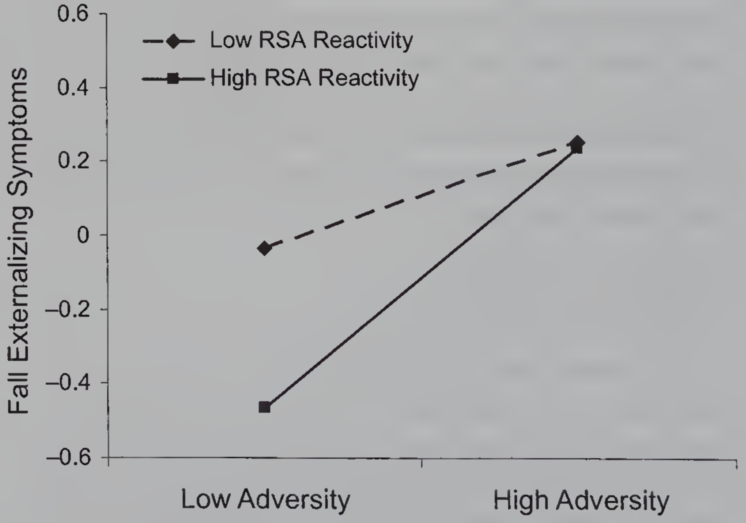


Figure 1. Fall externalizing symptoms as a function of adversity exposure and RSA reactivity.
Note. Low and high adversity values are graphed at 1 *SD* below and above the mean, respectively, whereas low and high RSA reactivity are graphed at 1 *SD* above and below the residual score mean, respectively. RSA = respiratory sinus arrhythmia.

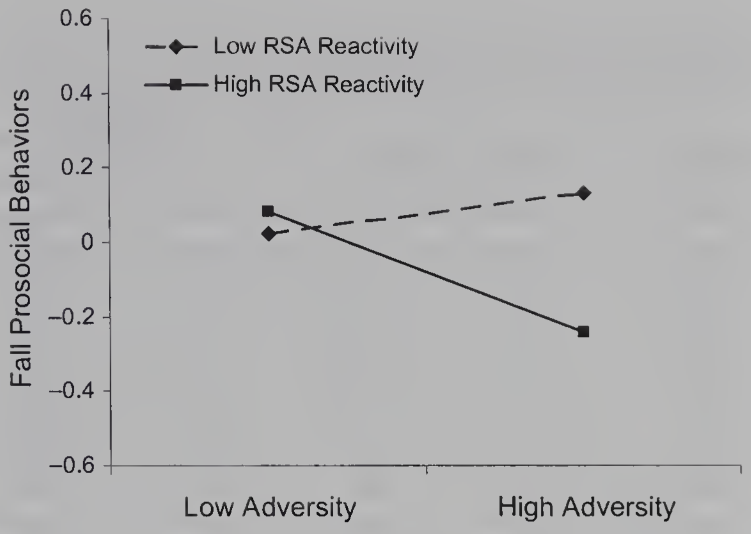


Figure 2. Fall prosocial behaviors as a function of adversity exposure and RSA reactivity.
Note. Low and high adversity values are graphed at 1 *SD* below and above the mean, respectively, whereas low and high RSA reactivity are graphed at 1 *SD* above and below the residual score mean, respectively. RSA = respiratory sinus arrhythmia.

decrease in academic competence in high-reactive children ($b = -.12$, $p < .05$). In the context of high adversity, low-reactive children showed higher levels of improvement in academic competence than high-reactive children (see Figure 4). However, in the context of low adversity, the reverse was true, with high-reactive children showing higher levels of improvement in academic competence than their low-reactive peers.

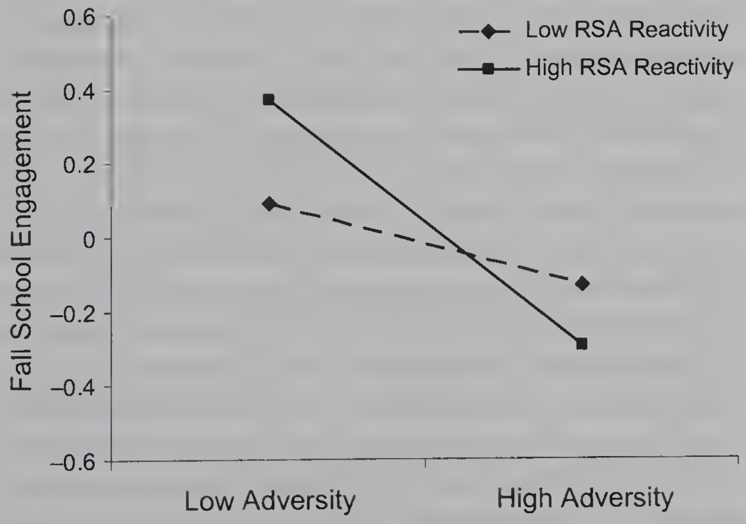


Figure 3. Fall school engagement as a function of adversity exposure and RSA reactivity.
Note. Low and high adversity values are graphed at 1 *SD* below and above the mean, respectively, whereas low and high RSA reactivity are graphed at 1 *SD* above and below the residual score mean, respectively. RSA = respiratory sinus arrhythmia.

The interaction effects of reactivity and adversity on different domains of adaptation did not vary across children's sex, as the three-way interaction effects between sex, reactivity, and adversity were not significant (see Table 4). However, a significant two-way interaction revealed that the main effect of RSA reactivity on 6-month change in academic competence varied for boys and girls. The test of simple slopes indicated a crossover effect such that high reactivity was related to academic improvement in boys ($b = -.06$, $p > .05$), whereas low reactivity was related to academic improvement in girls ($b = .10$, $p > .05$); however, neither slope reached statistical significance.

Cortisol reactivity. Results of regression analyses examining the effects of RSA reactivity and adversity on adaptation are presented in Table 5. High cortisol reactivity significantly predicted higher levels of externalizing and lower levels of school engagement and academic competence in fall. Adversity effects were the same as in regression analyses focused on RSA reactivity. However, these main effects are qualified by the significant interactions.

The interaction between cortisol reactivity and adversity significantly predicted fall prosocial behavior. According to tests of simple slopes, adversity exposure was significantly related to fall prosocial behaviors for high-reactive children ($b = -.16$, $p < .05$) but not in low-reactive children ($b = .06$, $p > .05$). In comparison to low-reactive peers, high-reactive children showed lower levels

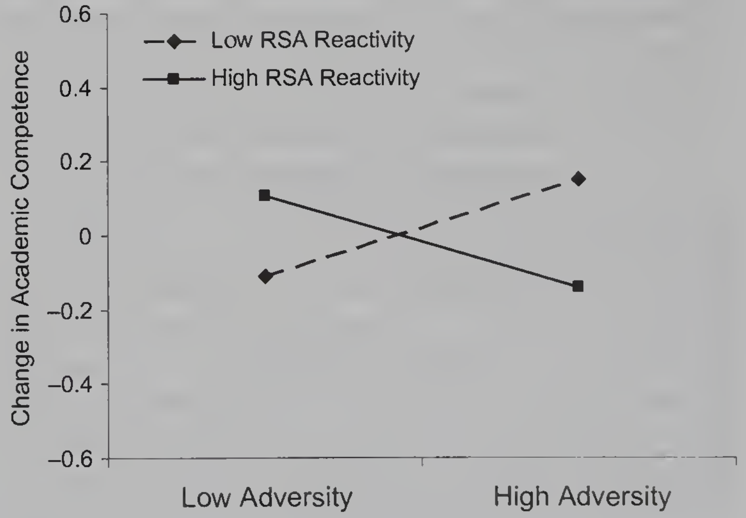


Figure 4. Change in academic competence as a function of adversity exposure and RSA reactivity.
Note. Low and high adversity values are graphed at 1 *SD* below and above the mean, respectively, whereas low and high RSA reactivity are graphed at 1 *SD* above and below the residual score mean, respectively. RSA = respiratory sinus arrhythmia.

Table 5

Standardized Regression Coefficients for Models Testing the Effects of Cortisol Reactivity and Adversity on Adaptation

	Externalizing		Prosocial behavior		School engagement		Academic competence	
	Fall	Spring	Fall	Spring	Fall	Spring	Fall	Spring
Fall adaptation		.70***		.65***		.54***		.75***
Sex	−.23***	−.11**	.32***	.15***	.17**	.05	.13*	.01
CORT	.14**	.00	−.02	−.05	−.13*	−.08	−.11*	−.02
ADV	.25***	.09*	−.05	−.09*	−.24***	−.10*	.00	.03
CORT × ADV	.05	.04	−.13*	−.05	−.04	−.06	−.08	.03
ADV × Sex	−.04	−.07*	.02	.03	.11*	.01	−.08	.00
CORT × Sex	−.12*	.04	.02	−.02	.07	.03	.07	.00
CORT × ADV × Sex	−.03	−.01	.09	−.03	.02	−.04	−.06	.00
Total R^2	.18***	.59***	.14***	.56***	.14***	.39***	.06**	.57***

Note. ACA = academic competence; ADV = adversity; CORT = cortisol reactivity; EXT = externalizing behavior problems; PRO = prosocial behavior; SE = school engagement.
$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

of fall prosocial behavior in the context of high adversity but in the context of low adversity showed slightly higher levels of prosocial behavior (see Figure 5).

The interaction effect of cortisol reactivity and adversity did not vary across children's sex, as indicated by nonsignificant three-way interactions (see Table 5). However, significant two-way interactions between sex and adversity emerged for fall school engagement and change in externalizing from fall to spring. Tests of simple slopes indicate that adversity exposure was significantly related to lower school engagement in boys ($b = .34$, $p < .001$) but not in girls ($b = -.13$, $p > .05$). Likewise, adversity exposure was related to an increase in externalizing symptoms in boys ($b = .16$, $p < .01$) but not in girls ($b = .01$, $p > .05$). In addition, the main effect of cortisol reactivity on fall externalizing differed for boys and girls, as indicated by a significant two-way interaction between reactivity and sex in Table 5. Cortisol reactivity was related to higher levels of externalizing in boys ($b = .31$, $p < .001$) but not in girls ($b = .03$, $p > .05$).

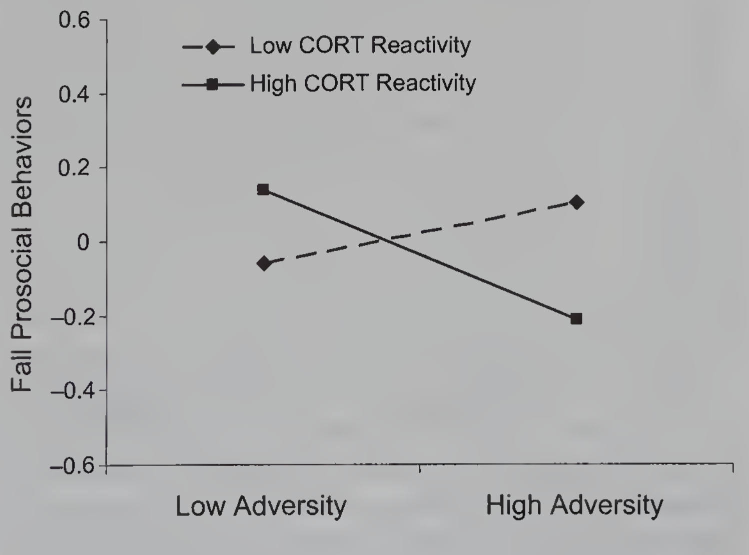


Figure 5. Prosocial behaviors as a function of adversity exposure and cortisol reactivity.
Note. Low and high adversity and cortisol reactivity values are graphed at 1 *SD* below and above the mean, respectively.

Discussion

The main goal of this study was to examine how stress reactivity and family adversity influence socioemotional behavior and school readiness skills in children attending kindergarten. Several significant main effects emerged. Consistent with the broad literature on the effects of risk and adversity (e.g., Luthar, 2006), family adversity was associated with concurrent externalizing symptoms and lower school engagement, and with increases in externalizing symptoms and decreases in prosocial behaviors and school engagement over the kindergarten year. However, the effect of adversity on fall school engagement and change in externalizing symptoms was significant only for boys. Adversity did not have a main effect on early kindergarten indices of academic competence. This discrepancy with previous research could be due to the fact that the majority of children in this study come from highly educated families, which may support academic success despite adversity exposure.

Low RSA reactivity was associated with higher levels of externalizing symptoms in the fall, corroborating previous findings of low stress reactivity and underarousal in children with externalizing symptoms (Boyce et al., 2001; Calkins et al., 2007; Mezzacappa et al., 1997; Pine et al., 1996; Pine et al., 1998) and the idea that low RSA reactivity is an index of emotional dysregulation and lability (Beauchaine, 2001). On the other hand, high cortisol reactivity was associated with high levels of externalizing behaviors and low levels of school engagement and academic competence. These findings are consistent with studies that have linked elevated cortisol in response to classroom challenges to poor self-regulation, impulsivity, peer rejection, and more solitary, negative behaviors (Dettling et al., 1999; Dettling et al., 2000; Gunnar et al., 1997; Gunnar et al., 2003), and the notion that hyperresponsivity of the HPA axis signals a risk for maladaptation (Gunnar & Vazquez, 2006). However, these main effects and the absence of others in these models should be interpreted with caution, as they were conditioned by significant interactions. For example, the effect of RSA reactivity on change in academic competence significantly varied across sex, with high reactivity being promotive for boys but a risk factor for girls. In addition, high cortisol reactivity was related to higher levels of fall externalizing symptoms only in boys. These findings suggest that children's sex may be an important factor to consider when examining the effects of stress reactivity on adaptation.

The study's most novel and salient findings emerged when adversity and stress reactivity were considered together, as components of interactions between environmental exposures and measures of biological sensitivity. Stress reactivity moderated the negative effect of family adversity across various domains of adaptation. Overall, the findings are consistent with the stress diathesis hypothesis that high-reactive children show worse adaptive functioning in the context of high adversity. Indeed, such children generally evinced the lowest levels of adaptive functioning of the entire study sample.

However, equally reactive children in settings of low adversity showed the highest levels of adaptation, levels even higher than those of their less reactive counterparts. Specifically, in the context of low family adversity, children who showed high RSA reactivity in response to challenges had the lowest levels of externalizing symptoms and the highest levels of prosocial behaviors and school engagement. Although adaptation showed significant stability from fall to spring, high-reactive children showed improvement in academic competence in the context of low adversity and a decline in competence in the context of high adversity, whereas the inverse was true for low reactive children. Similarly, children who showed high cortisol reactivity to the challenge protocol had the highest levels of prosocial behaviors in the context of low adversity. Further, children exhibiting low RSA reactivity in response to challenges were fully or partially buffered against the harmful effects of adversity on externalizing symptoms, prosocial behavior, and school engagement. Likewise, among children who showed low cortisol reactivity, levels of prosocial behaviors did not significantly change across different levels of adversity.

These findings support the biological sensitivity to context theory advanced by Boyce and colleagues (Boyce 2007; Boyce & Ellis, 2005) and the concept of differential susceptibility to environmental influences proposed by Belsky and colleagues (Belsky, 2005; Belsky et al., 2007). This study illustrates that high reactivity is not merely a pathogenic, risk-amplifying response to adversity but can also promote adaptive functioning. Corroborating Boyce and colleagues' theoretical perspective, children exhibiting high levels of biological sensitivity to context, as indexed by high autonomic and adrenocortical reactivity, were more susceptible to environmental influences in the context of both low and high family adversity. Thus, biologically sensitive children showed the highest levels of symptoms in the context of high family adversity but the highest levels of competence in the context of low family adversity. However, a lack of family adversity does not necessarily imply the presence of a nurturing family environment. Thus, future studies will need to further examine the role of heightened biological sensitivity to context across both stressful, health-undermining and supportive, health-enhancing contexts.

Together, these findings provide strong support for reframing stress reactivity as a biological sensitivity to context. First, the findings are relatively robust, especially for RSA reactivity, which emerged as a significant moderator of adversity effects across all four domains of adaptation. The evidence was parallel, though less robust, for cortisol reactivity, as the interaction effect was significant only for one domain. Second, each observed interaction was consistent with the biological sensitivity to context theory; high reactivity was promotive in the context of low adversity but a risk factor in the context of high adversity. Third, the interaction effect was found for both positive and negative indices of adaptation. Past studies have rarely examined how interactive processes apply to multiple domains of functioning and have often focused primarily on measures of psychopathology. Fourth, these highly consistent interaction effects were found in a normative, community sample of typically developing children. If stress reactivity moderated the effects of adversity on development in a community sample of children, who on average do not face extreme disadvantage or show clinical levels of problem behavior, the interactive effects are likely to be larger at the extreme ends of such distributions.

Finally, the reported interaction effects complement recent work examining individual differences in markers of behavioral and genetic susceptibility to environmental influences. The current findings are consistent with research on behavioral reactivity indicating that children with high levels of negative affectivity are particularly susceptible to both negative and positive experiences (Belsky, 2005; Belsky et al., 2007; Klein Velderman et al., 2006). In addition, the current findings parallel recent research showing that genetic polymorphisms can moderate the effects on adaptive functioning of family adversity, unresolved trauma, or abusive parenting (Bakermans-Kranenburg & van IJzendoorn, 2007; Caspi et al., 2002; Taylor et al., 2006; van IJzendoorn & Bakermans-Kranenburg, 2006) and resonate with the observations of Rutter and colleagues on the previously overlooked protective effects of "risk-engendering" genetic variants (Rutter, Moffitt, & Caspi, 2006). Future studies should examine whether these behavioral, physiological, and genetic markers of susceptibility to contextual factors represent the same phenomena expressed at different levels of assessment or whether they represent different types of susceptibility that may have a cumulative effect on development.

For the purposes of future studies, it is important to note that the nature of the reported interaction effects may depend on several dimensions depicted in the conceptual model of Figure 6. Single studies may be capable of capturing only a portion of the underlying, paradigmatic interaction shown in Figure 6, as the "observational window" shifts from study to study depending on sampling characteristics, the outcomes of the interest, the measure of reactivity, and the type of challenge. For example, the observational window may shift to the right in studies of highly disadvantaged child populations or upward in studies of children with severe behavior problems. Similarly, the slopes of the two component lines may be determined by the researchers' ability to capture the full range of stress reactivity responses and thus distinguish between high- and low-reactive children. It is critical, therefore, that developmental researchers examine and compare the interactive effects of stress reactivity and context across a range of outcomes and environmental influences and that the full scope of existing literature be taken into account in the interpretation of such effects.

The relations between stress reactivity and context are also likely to be dynamic and evolving over time and the present study provides only a snapshot into a potentially complex cycle of bidirectional and reverberatory influences. Boyce and colleagues (Boyce, 2006, 2007; Boyce et al., 2005) have argued, for example, that stress reactivity may not only moderate effects of early experience but may itself *be influenced by* the quality and character of early experience. While our analyses addressed stress reactivity as a moderator of adversity, given the developmental stage of the participants and the nature of the study data, we are mindful that adversity exposure could also play a role in shaping individuals' stress reactivity. From the perspective of evolutionary biology (Ellis, Jackson, & Boyce, 2006), children growing up in supportive and nurturing families might be expected to develop high levels of biological sensitivity to context in order to take greater advantage of the positive and stimulating features of their environments. As a result of both high sensitivity and protective environments, these children would show high levels of competence and low rates of mental and health problems. On the other hand, children

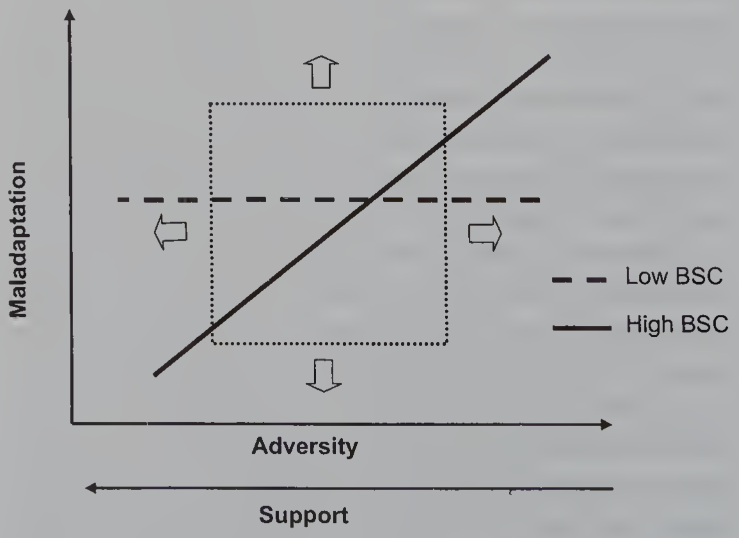


Figure 6. Conceptual model of the interactive effects between biological sensitivity to context (BSC) and environmental influences.

exposed to high levels of early risk and adversity may also develop high biological sensitivity to context, as a means of sustaining vigilance for environmental threats and hazards. Although such vigilance may be adaptive in the short run, over a longer period of time it could augment children's vulnerability to the deleterious effects of adversity. It is not surprising, then, that these children tend to have higher rates of mental and health problems. On the other hand, children who develop low biological sensitivity to context seem to be less affected by both positive and negative environmental influences. These children may consequently demonstrate resilience—adaptation that is better than expected given their adversity exposures—or vulnerability—low levels of adaptation despite growing up in environments abundant with resources and support (Luthar, 2006; Masten & Obradović, 2006). Future studies are needed to examine the processes by which early experiences shape children's biological stress responses.

Strengths, Limitations, and Future Directions

This study had several unique strengths as well as notable limitations that highlight important directions for future improvements. First, the study was based on a large, ethnically diverse sample of kindergarten children; however, a majority of children came from families with relatively high socioeconomic status, as indicated by income and parents' education. Second, the study was unique in that it employed a carefully controlled measure of autonomic reactivity that took into account the possible effects of motor activity and engagement demands elicited by challenge tasks. However, only one third of the sample had elevated cortisol in response to stress reactivity protocol, which may have contributed to less robust findings for cortisol reactivity. Future studies should attempt to raise the responder rate, as a lack of response to the current protocol does not necessarily indicate a low reactivity to more intense, real-life stressors. Third, we included the various sources of family adversity as an index of early experience. However, future researchers should explicitly assess positive environmental influences in children's lives by including measures of social support and resources. Another strength of the current study was the use of multiple informants. While it may not always be feasible to use multiple informants when assessing family adversity, given the young age of the participants and the nature of the construct, future studies should, where possible, triangulate the child, parent, and teacher reports of adaptation, thus minimizing informant variance. Fourth, the study was unique in its effort to compare how both indices of autonomic and adrenocortical reactivity interact with early family adversity to predict adaptation. In addition to examining how indices of different types of stress reactivity affect children's behaviors and skills, future research should examine how different stress response systems interact to produce unique stress reactivity profiles, which may be related to different adaptive patterns across adversity exposure. As discussed earlier, researchers might use person-focused analytic approaches to examine whether the same children express analogous behavioral, physiological, and genetic sensitivity to contextual factors. Finally, future research should examine transactional relations between stress reactivity and adversity across longer periods of development, as adaptation in this study showed high stability over the 6-month period.

Conclusion

Given the pervasive and long-lasting effects that adversity can have on adaptation (Luthar, 2006; Obradović et al., in press; Sameroff, 2006), it is important to elucidate the processes that buffer or exacerbate these effects. This study represents a rare attempt to examine complex interactions between biological and environmental factors. The findings indicate that children's biological sensitivity to social context played an important role in moderating the effects of early experiences of family adversity on positive and negative indices of adaptive functioning. We believe that studies such as the one reported here can advance future prevention and intervention efforts, to address more holistically the social disparities in children's competence and psychopathology.

References

Ablow, J. C., Measelle, J. R., & the MacArthur Working Group on Outcome Assessment. (2003). *Manual for the Berkeley Puppet Interview: Symptomatology, social, and academic modules (BPI 1.0)*. MacArthur Foundation Research Network on Psychopathology and Development. Pittsburgh, PA: University of Pittsburgh Press.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Alkon, A., Goldstein, L., Smider, N., Essex, M., Boyce, W., & Kupfer, D. (2003). Developmental and contextual influences on autonomic reactivity in young children. *Developmental Psychobiology, 42*, 64–78.

Armstrong, J. M., Goldstein, L. H., & the MacArthur Working Group on Outcome Assessment. (2003). *Manual for the MacArthur Health and Behavior Questionnaire (HBQ 1.0)*. MacArthur Foundation Research Network on Psychopathology and Development (David J. Kupfer, Chair). Pittsburgh, PA: University of Pittsburgh.

Bakermans-Kranenburg, M. J., & Van IJzendoorn, M. H. (2007). Genetic vulnerability or differential susceptibility in child development: The case of attachment. *Journal of Child Psychology and Psychiatry, 48*, 1160–1173.

Bar-Haim, Y., Marshall, P., & Fox, N. (2000). Developmental changes in heart period and high-frequency heart period variability from 4 months to 4 years of age. *Developmental Psychobiology, 37*, 44–56.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173–1182.

Bazhenova, O. V., & Porges, S. W. (1997). The integrative neurobiology of affiliation. *Annals of the New York Academy of Sciences, 807*, 469–471.

Beauchaine, T. (2001). Vagal tone, development, and Gray's motivational theory: Toward an integrated model of autonomic nervous system functioning in psychopathology. *Development and Psychopathology, 13*, 183–214.

Beauchaine, T., Gatzke-Kopp, L., & Mead, H. (2007). Polyvagal theory and developmental psychopathology: Emotion dysregulation and conduct problems from preschool to adolescence. *Biological Psychology, 74*, 174–184.

Belsky, J. (2005). Differential susceptibility to rearing influence: An evolutionary hypothesis and some evidence. In B. Ellis & D. Bjorklund (Eds.), *Origins of the social mind: Evolutionary psychology and child development* (pp. 139–163). New York: Guilford.

Belsky, J., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2007). For better and for worse: Differential susceptibility to environmental influences. *Current Directions in Psychological Science, 16*, 300–304.

Belsky, J., Hsieh, K., & Crnic, K. (1998). Mothering, fathering, and infant negativity as antecedents of boys' externalizing problems and inhibition at age 3: Differential susceptibility to rearing influence? *Development and Psychopathology, 10*, 301–319.

Bernardi, L., Wdowczyk-Szulc, J., Valenti, C., Castoldi, S., Passino, C., Spadacini, G., et al. (2000). Effects of controlled breathing, mental activity and mental stress with or without verbalization on heart rate variability. *Journal of the American College of Cardiology, 35*, 1462–1469.

Berntson, G., Bigger, J., Eckberg, D., Grossman, P., Kaufman, P., Malik, M., et al. (1997). Heart rate variability: Origins, methods, and interpretive caveats. *Psychophysiology, 34*, 623–648.

Berntson, G., Cacioppo, J., & Fieldstone, A. (1996). Illusions, arithmetic, and the bidirectional modulation of vagal control of the heart. *Biological Psychology, 46*, 1–17.

Berntson, G., Cacioppo, J., & Quigley, K. (1993). Respiratory sinus arrhythmia: Autonomic origins, physiological mechanisms, and psychophysiological implications. *Psychophysiology, 30*, 183–196.

Berntson, G. G., Quigley, K. S., Jang, J., & Boysen, S. T. (1990). An approach to artifact identification: Application to heart period data. *Psychophysiology, 27*, 586–598.

Blair, C. (2002). Early intervention for low birth weight preterm infants: The role of negative emotionality in the specification of effects. *Development and Psychopathology, 14*, 311–332.

Blair, C., Granger, D., & Razza, R. P. (2005). Cortisol reactivity is positively related to executive function in preschool children attending head start. *Child Development, 76*, 554–567.

Block, J. H. (1965). *The child-rearing practices report*. Unpublished manuscript, University of California, Berkeley.

Boyce, W. T. (1996). Biobehavioral reactivity and injuries in children and adolescents. In M. H. Bornstein & J. Genevro (Eds.), *Child development and behavioral pediatrics: Toward understanding children and health* (pp. 35–58). Mahwah, NJ: Erlbaum.

Boyce, W. T. (2006). Biology and context: Symphonic causation and the origins of childhood psychopathology. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology: Developmental neuroscience* (Vol. II, pp. 797–817). Hoboken, NJ: Wiley.

Boyce, W. T. (2007). A biology of misfortune: Stress reactivity, social context, and the ontogeny of psychopathology in early life. In A. Masten (Ed.), *Multilevel dynamics in developmental psychopathology: Pathways to the future* (34th ed., pp. 45–82). Minneapolis: University of Minnesota.

Boyce, W. T., Chesney, M., Alkon-Leonard, A., Tschann, J., Adams, S., Chesterman, B., et al. (1995). Psychobiologic reactivity to stress and childhood respiratory illnesses: Results of two prospective studies. *Psychosomatic Medicine, 57*, 411–422.

Boyce, W. T., & Ellis, B. J. (2005). Biological sensitivity to context: I. An evolutionary–developmental theory of the origins and functions of stress reactivity. *Development and Psychopathology, 17*, 271–301.

Boyce, W. T., Essex, M. J., Alkon, A., Goldsmith, H. H., Kraemer, H. C., & Kupfer, D. J. (2006). Early father involvement moderates biobehavioral susceptibility to mental health problems in middle childhood. *Journal of the American Academy of Child & Adolescent Psychiatry, 45*, 1510–1520.

Boyce, W. T., Quas, J., Alkon, A., Smider, N., Essex, M., & Kupfer, D. J. (2001). Autonomic reactivity and psychopathology in middle childhood. *British Journal of Psychiatry, 179*, 144–150.

Burchinal, M., Roberts, J. E., Hooper, S., & Zeisel, S. A. (2000). Cumulative risk and early cognitive development:

A comparison of statistical risk models. *Developmental Psychology, 36*, 793–807.

Bush, N. R., Alkon, A., Obradović, J., Stamperdahl, J., & Boyce, W. T. (2009). *Deconstructing psychobiological reactivity: The contributions of psychomotor activity and stress response in five year old children*. Manuscript submitted for publication.

Cacioppo, J., Uchino, B., & Berntson, G. (1994). Individual differences in the autonomic origins of heart rate reactivity: The psychometrics of respiratory sinus arrhythmia and preejection period. *Psychophysiology, 31*, 412–419.

Calkins, S. D. (1997). Cardiac vagal tone indices of temperamental reactivity and behavioral regulation in young children. *Developmental Psychobiology, 31*, 125–135.

Calkins, S. D., Graziano, P. A., & Keane, S. P. (2007). Cardiac vagal regulation differentiates among children at risk for behavior problems. *Biological Psychology, 74*, 144–153.

Carlson, R. (1985). Gesell School Readiness Test. In D. Keyser & R. Sweetland (Eds.), *Test critiques*. Kansas City, KS: Test Corporation of America.

Carlson, M., & Earls, F. (1997). Psychological and neuroendocrinological sequelae of early social deprivation in institutionalized children in Romania. *Annals of the New York Academy of Sciences, 807*, 419–428.

Caspi, A., McClay, J., Moffitt, T., Mill, J., Martin, J., Craig, I., et al. (2002, August 2). Role of genotype in the cycle of violence in maltreated children. *Science, 297*, 851–854.

Crowell, S., Beauchaine, T. P., McCauley, E., Smith, C., Stevens, A. L., & Sylvers, P. (2005). Psychological, autonomic, and serotonergic correlates of parasuicidal behavior in adolescent girls. *Development and Psychopathology, 17*, 1105–1127.

Cummings, E. M., & Davies, P. T. (2002). Effects of marital conflict on children: Recent advances and emerging themes in process-oriented research. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 43*, 31–63.

Cummings, E. M., El-Sheikh, M., Kouros, C. D., & Keller, P. S. (2007). Children's skin conductance reactivity as a mechanism of risk in the context of parental depressive symptoms. *Journal of Child Psychology and Psychiatry, 48*, 436–445.

Curtis, W. J., & Cicchetti, D. (2003). Moving research on resilience into the 21st century: Theoretical and methodological considerations in examining the biological contributors to resilience. *Development and Psychopathology, 15*, 773–810.

Davis, E. P., Bruce, J., & Gunnar, M. R. (2002). The anterior attention network: Associations with temperament neuroendocrine activity in 6-year-old children. *Developmental Psychobiology, 40*, 43–56.

Davis, E. P., Donzella, B., Krueger, W. K., & Gunnar, M. (1999). The start of a new school year: Individual differences in salivary cortisol response in relation to child temperament. *Developmental Psychobiology, 35*, 188–196.

Deater-Deckard, K., & Dodge, K. (1997). Spare the rod, spoil the authors: Emerging themes in research on parenting. *Psychological Inquiry, 8*, 230–235.

Deković, M., Janssens, J. M. A. M., & Gerris, J. R. M. (1991). Factor structure and construct validity of the Block Child Rearing Practice Report (CRPR). *Psychological Assessment: Journal of Consulting & Clinical Psychology, 3*, 182–187.

Dettling, A. C., Gunnar, M. R., & Donzella, B. (1999). Cortisol levels of young children in full-day childcare centers: Relations with age and temperament. *Psychoneuroendocrinology, 24*, 519–536.

Dettling, A. C., Parker, S. W., Lane, S., Sebanc, A., & Gunnar, M. R. (2000). Quality of care and temperament determine changes in cortisol concentrations over the day for young children in childcare. *Psychoneuroendocrinology, 25*, 819–836.

Dietrich, A., Riese, H., Sondeijker, F. E. P. L., Graces-Lord, K., van Roon, A. M., Ormel, J., et al. (2007). Externalizing and internalizing problems in relation to autonomic function: A population-based study in preadolescents. *Journal of American Academy of Child and Adolescent Psychiatry, 46*, 378–386.

Doussard-Roosevelt, J. A., Montogomery, L. A., & Porges, S. W. (2003). Short-term stability of physiological measures in kindergarten children: Respiratory sinus arrhythmia, heart period, and cortisol. *Developmental Psychobiology, 43*, 230–242.

Eisenberg, N., Fabes, R., Bustamante, D., Mathy, R., Miller, P., & Lindholm, E. (1988). Differentiation of vicariously induced emotional reactions in children. *Developmental Psychology, 24*, 237–246.

Eisenberg, N., Fabes, R. A., Murphy, B., Maszk, P., Smith, M., & Karbon, M. (1995). The role of emotionality and regulation in children's social functioning: A longitudinal study. *Child Development, 66*, 1360–1384.

Ellis, B. J., Essex, M. J., & Boyce, W. T. (2005). Biological sensitivity to context: II. Empirical explorations of an evolutionary–developmental theory. *Development and Psychopathology, 17*, 303–328.

Ellis, B. J., Jackson, J. J., & Boyce, W. T. (2006). The stress response systems: Universality and adaptive individual differences. *Developmental Review, 26*, 175–212.

El-Sheikh, M. (2005). The role of emotional responses and physiological reactivity in the marital conflict-child functioning link. *Journal of Child Psychology and Psychiatry, 46*, 1191–1199.

El-Sheikh, M., Harger, J., & Whitson, S. M. (2001). Exposure to interparental conflict and children's adjustment and physical health: The moderating role of vagal tone. *Child Development, 72*, 1617–1636.

El-Sheikh, M., & Whitson, S. A. (2006). Longitudinal relations between marital conflict and child adjustment: Vagal regulation as a protective factor. *Journal of Family Psychology, 20*, 30–39.

Essex, M. J., Klein, M. H., Cho, E., & Kalin, N. H. (2002). Maternal stress beginning in infancy may sensitize children to later stress exposure: Effects on cortisol and behavior. *Biological Psychiatry, 52*, 776–784.

Fabes, R. A., Eisenberg, N., & Eisenbud, L. (1993). Behavioral and physiological correlates of children's reactions to others in distress. *Developmental Psychology, 29*, 655–663.

Field, T., Land, C., Martinez, A., Yando, R., Pickens, J., & Bendell, D. (1996). Preschool follow-up of infants of dysphoric mothers. *Journal of Clinical Child Psychology, 25*, 272–279.

Fox, N. A., & Field, T. M. (1989). Individual differences in preschool entry behavior. *Journal of Applied Developmental Psychology, 10*, 527–540.

Goodyer, I. M., Herbert, J., & Althan, P. (1998). Adrenal steroid secretion and major depression in 8- to 16-year olds, III. Influence of cortisol/DHEA ratio at presentation on subsequent rates of disappointing life events and persistent major depression. *Psychological Medicine, 28*, 265–273.

Granger, D. A., Weisz, J. R., & Kauneckis, D. (1994). Neuroendocrine reactivity, internalizing behavior problems, and control-related cognitions in clinic-referred children and adolescents. *Journal of Abnormal Psychology, 103*, 267–276.

Granger, D. A., Weisz, J. R., McCracken, J. T., Ikeda, S. C., & Douglas, P. (1996). Reciprocal influences among adrenocortical activation, psychosocial processes, and the behavioral adjustment of clinic-referred children. *Child Development, 67*, 3250–3262.

Gunnar, M. R., Sebanc, A. M., Tout, K., Donzella, B., & van Dulmen, M. H. (2003). Peer rejection, temperament, and cortisol activity in preschoolers. *Developmental Psychobiology, 43*, 346–358.

Gunnar, M. R., Tout, K., de Haan, M., Pierce, S., & Stansbury, K. (1997). Temperament, social competence, and adrenocortical activity in preschoolers. *Developmental Psychobiology, 31*, 65–85.

Gunnar, M. R., & Vazquez, D. (2006). Stress neurobiology and developmental psychopathology. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology* (Vol. 2, pp. 533–577). Hoboken, NJ: Wiley.

Halberstadt, A. G. (1986). Family socialization of emotional expression and nonverbal communication styles and skills. *Journal of Personality and Social Psychology, 51*, 827–836.

Hardie, T., Moss, H., Vanyukov, M., Yao, J., & Kirillovac, G. (2002). Does adverse family environment or sex matter in the salivary cortisol responses to anticipatory stress? *Psychiatry Research, 111*, 121–131.

Johnson, P. L., & O'Leary, K. D. (1987). Parental behavior patterns and conduct problems in girls. *Journal of Abnormal Child Psychology, 15*, 573–581.

Kagan, J., & Snidman, N. (1991). Temperamental factors in human development. *American Psychologist, 46*, 856–862.

Kamarck, T. W., & Lovallo, W. R. (2003). Cardiovascular reactivity to psychological challege: Conceptual and measurement considerations. *Psychosomatic Medicine, 65*, 9–21.

Katz, L. F., & Gottman, J. M. (1995). Vagal tone protects children from marital conflict. *Development and Psychopathology, 7*, 83–92.

Katz, L. F., & Gottman, J. M. (1997). Buffering children from marital conflict and dissolution. *Journal of Clinical Child Psychology, 26*, 157–171.

Kaufman, A., & Kaufman, N. (1983). *Kaufman Assessment Battery for Children*. Circle Pines, MN: American Guidance Service.

King, J., Barkley, R. A., & Barrett, S. (1998). Attention-deficit hyperactivity disorder and the stress response. *Biological Psychiatry, 44*, 72–74.

Klein Velderman, M., Bakermans-Kranenburg, M. J., Juffer, F., & van IJzendoorn, M. H. (2006). Effects of attachment-based interventions on maternal sensitivity and infant attachment: Differential susceptibility of highly reactive infants. *Journal of Family Psychology, 20*, 266–274.

Klimes-Dougan, B., Hastings, P. D., Granger, D. A., Usher, B. A., & Zahn-Waxler, C. (2001). Adrenocortical activity in at-risk and normally developing adolescents: Individual differences in salivary cortisol basal levels, diurnal variation, and responses to social challenges. *Development and Psychopathology, 13*, 695–719.

Kraemer, H. C., Measelle, J. R., Ablow, J. C., Essex, M. J., Boyce, W. T., & Kupfer, D. J. (2003). A new approach to integrating data from multiple informants in psychiatric assessment and research: Mixing and matching contexts and perspectives. *American Journal of Psychiatry, 160*, 1566–1577.

Kudielka, B. M., Schommer, N. C., Hellhammer, D. H., & Kirschbaum, C. (2004). Acute HPA axis responses, heart rate, and mood changes to psychosocial stress (TSST) in humans at different times of day. *Psychoneuroendocrinology, 29*, 983–992.

Leary, A., & Katz, L. F. (2004). Coparenting, family-level processes, and peer outcomes: The moderating role of vagal tone. *Development and Psychopathology, 16*, 593–608.

Luthar, S. S. (2006). Resilience in development: A synthesis of research across five decades. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology* (2nd ed., Vol. 3, pp. 739–795). Hoboken, NJ: Wiley.

Manuck, S., Kasprowicz, A., & Muldoon, M. (1990). Behaviorally evoked cardiovascular reactivity and hypertension: Conceptual issues and potential associations. *Annals of Behavioral Medicine, 12*, 17–29.

Masharani, U., Shiboski, S., Eisner, M. D., Katz, P. P., Janson, S. L., Granger, D. A., et al. (2005). Impact of exogenous glucocorticoid use on salivary cortisol measurements among adults with asthma and rhinitis. *Psychoneuroendocrinology, 30*, 744–752.

Masten, A. S., & Obradović, J. (2006). Competence and resilience in development. *Annals of the New York Academy of Sciences, 1094*, 13–27.

Masten, A. S., & Shaffer, A. (2006). How families matter in child development: Reflections from research on risk and resilience. In A. Clarke-Stewart & J. Dunn (Eds.), *Families count: Effects on child and adolescent development* (pp. 5–25). New York: Cambridge University Press.

McBurnett, K. M., Lahey, B. B., Frick, P. J., Risch, C., Loeber, R., Hart, E. L., et al. (1991). Anxiety, inhibition, and conduct disorder in children: II. Relation to salivary cortisol. *Journal of the American Academy of Child and Adolescent Psychiatry, 30,* 192–196.

McBurnett, K. M., Lahey, B. B., Rathouz, P. J., & Loeber, R. (2000). Low salivary cortisol and persistent aggression in boys referred for disruptive behavior. *Archives of General Psychiatry, 57,* 38–43.

Mezzacappa, E., Tremblay, R. E., Kindlon, D., Saul, J. P., Arseneault, L., Seguin, J., et al. (1997). Anxiety, antisocial behavior and heart rate regulation in adolescent males. *Journal of Child Psychology and Psychiatry, 38,* 457–469.

Moss, H. B., Vanyukov, M. M., & Martin, C. S. (1995). Salivary cortisol responses and the risk for substance abuse in prepubertal boys. *Biological Psychiatry, 38,* 547–555.

Obradović, J., Shaffer, A., & Masten, A. S. (in press). Adversity and risk in developmental psychopathology: Progress and future directions. In L. C. Mayes & M. Lewis (Eds.), *The environment of human development: A handbook of theory and measurement* New York: Cambridge University Press.

Pine, D. S., Wasserman, G., Coplan, J., Fried, J., Sloan, R., Myers, M., et al. (1996). Serotonergic and cardiac correlates of aggression in children. *Annals of the New York Academy of Sciences, 794,* 391–393.

Pine, D. S., Wasserman, G. A., Miller, L., Coplan, J. D., Bagiella, E., Kovelenku, P., et al. (1998). Heart period variability and psychopathology in urban boys at risk for delinquency. *Psychophysiology, 35,* 521–529.

Porges, S. W. (2001). The polyvagal theory: Phylogenetic substrates of a social nervous system. *International Journal of Psychophysiology, 42,* 123–146.

Porges, S. W. (2003). The polyvagal theory: Phylogenetic contributions to social behavior. *Physiology and Behavior, 79,* 503–513.

Porges, S. W. (2007). The polyvagal perspective. *Biological Psychology, 74,* 174–184.

Porges, S. W., Heilman, K. J., Bazhenova, O., Bal, E., Doussard-Roosevelt, J. A., & Koledin, M. (2007). Does motor activity during psychophysiological paradigms confound the quantification and interpretation of heart rate and heart rate variability measures in young children? *Developmental Psychobiology, 49,* 485–494.

Porter, B., & O'Leary, K. D. (1980). Marital discord and childhood behavior problems. *Journal of Abnormal Child Psychology, 8,* 287–295.

Radloff, L. S. (1977). The CES–D scale: A self report depression scale for research in the general population. *Applied Psychological Measurement, 1,* 385–401.

Ramos, M. C., Guerin, D. W., Gottfried, A. W., Bathurst, K., & Oliver, P. H. (2005). Family conflict and children's behavior problems: The moderating role of child temperament. *Structural Equation Modeling, 12,* 278–298.

Rickel, A. U., & Biasatti, L. L. (1982). Modification of the Block Child Rearing Practice Report. *Journal of Clinical Psychology, 38,* 129–133.

Rudolph, C. D., Rudolph, A. M., Hostetter, M. K., Lister, G. L., & Siegel, N. J. (2003). *Rudolph's pediatrics* (21st ed.). New York: McGraw-Hill.

Rutter, M. (1983). Stress, coping, and development: Some issues and some questions. In N. Garmezy & M. Rutter (Eds.), *Stress, coping, and development in children* (pp. 1–41). New York: McGraw-Hill.

Rutter, M., Moffitt, T. E., & Caspi, A. (2006). Gene-environment interplay and psychopathology: Multiple varieties but real effects. *Journal of Child Psychology and Psychiatry, 47,* 226–261.

Sameroff, A. (2006). Identifying risk and protective factors for healthy child development. In A. Clark-Stewart & J. Dunn (Eds.), *Families count: Effects on child and adolescent development* (pp. 53–76). Cambridge, UK: Cambridge University Press.

Sapolsky, R. M., Romero, L. M., & Munck, A. U. (2000). How do glucocorticoids influence stress responses? Integrating permissive, suppressive, stimulatory, and preparative actions. *Endocrine Reviews, 21,* 55–89.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7,* 147–177.

Shannon, K. E., Beauchaine, T. P., Brenner, S. L., Neuhaus, E., & Gatzke-Kopp, L. (2007). Familial and temperamental predictors of resilience in children at risk for conduct disorder and depression. *Development and Psychopathology, 19,* 701–727.

Shirtcliff, E. A., Granger, D. A., Booth, A., & Johnson, D. (2005). Low salivary cortisol levels and externalizing behavior problems in youth. *Development and Psychopathology, 17,* 167–184.

Smider, N. A., Essex, M. J., Kalin, N. H., Buss, K. A., Klein, M. H., Davidson, R. J., et al. (2002). Salivary cortisol as a predictor of socioemotional adjustment during kindergarten: A prospective study. *Child Development, 73,* 75–92.

Spielberger, C. D. (1988). *Manual for the State-Trait Anger Expression Inventory (STAXI).* Odessa, FL: Psychological Assessment Resources.

Suess, P. E., Porges, S. W., & Plude, D. J. (1994). Cardiac vagal tone and sustained attention in school-age children. *Psychophysiology, 31,* 17–22.

Taylor, S. E., Way, B. M., Welch, W. T., Hilmert, C. J., Lehman, B. J., & Eisenberger, N. I. (2006). Early family environment, current adversity, the serotonin transporter polymorphism, and depressive symptomatology. *Biological Psychiatry, 60,* 671–676.

van IJzendoorn, M. H., & Bakermans-Kranenburg, M. J. (2006). DRD4 7-repeat polymorphism moderates the association between maternal unresolved loss or trauma and infant disorganization. *Attachment and Human Development, 8,* 291–307.

Child Development, January/February 2010, Volume 81, Number 1, Pages 290–305

Understanding Relations Among Early Family Environment, Cortisol Response, and Child Aggression via a Prevention Experiment

Colleen R. O'Neal, Laurie Miller Brotman, Keng-Yen Huang, Kathleen Kiely Gouley, Dimitra Kamboukos, and Esther J. Calzada
NYU Child Study Center, New York University School of Medicine

Daniel S. Pine
National Institute of Mental Health

This study examined relations among family environment, cortisol response, and behavior in the context of a randomized controlled trial with 92 children ($M = 48$ months) at risk for antisocial behavior. Previously, researchers reported an intervention effect on cortisol response in anticipation of a social challenge. The current study examined whether changes in cortisol response were related to later child aggression. Among lower warmth families, the intervention effect on aggression was largely mediated by the intervention effect on cortisol response. Although the intervention also resulted in significant benefits on child engaging behavior, cortisol response did not mediate this effect. These findings demonstrate meaningful associations between cortisol response and aggression among children at familial risk for antisocial behavior.

The way in which children respond to stress is an important aspect of development. Although experimental studies with animals have led to influential theoretical modeling of the role of human cortisol regulation in maladaptive outcomes (Champagne & Meaney, 2006; Francis, Caldji, Champagne, Plotsky, & Meaney, 1999; Levine, 1957; Liu et al., 1997; Meaney, 2001), only limited experimental work has been done with humans. In children, cross-sectional studies largely support a negative relation between cortisol response to a stressor and aggressive behavior (Blair, Granger, & Razza, 2005; Snoek, Van Goozen, Matthys, Buitelaar, & Van Engeland, 2004; van Goozen, Matthys, Cohen-Kettenis, Buitelaar, & van Engeland, 2000; van Goozen et al., 1998), although the exact nature of the relation has yet to be established (Dadds & Rhodes, 2008; Hart, Burock, London, Atkins, & Bonilla-Santiago, 2005; Rutter, 2008).

A rich literature demonstrates a relation between a blunted child cortisol response to stress and unresponsive or neglectful environments (Gunnar, Fisher, & the Early Experience, Stress and Prevention Network, 2006). However, as with the cortisol–behavior association, the nature of these cortisol–context relations is not well understood.

One model of the development of an abnormal stress response (e.g., Caldji, Diorio, & Meaney, 2000; Liu et al., 1997) suggests that mothers living under stressful conditions display abnormal stress responses, which may be transmitted directly to their offspring. Such maladaptive stress responsivity may lead to various negative outcomes, including aggression. Other models emphasize the complex manner in which genes and environment shape physiology and behavior through Gene × Environment interactions and Gene × Environment correlations (Rutter, Moffitt, & Caspi, 2006). For example, genes that shape maternal stress response could impact the environment by influencing

---

This research was supported in part by a NARSAD Young Investigator Award to the first author and a grant from the National Institute of Mental Health (NIMH R01 MH55188) to the second author. We acknowledge the research staff of the Institute for Prevention Science, Tom Cooper and staff of the Nathan Kline Institute for assistance with the processing of cortisol assays, and the families and children who participated. We would also like to acknowledge the substantial contributions that Drs. Rachel Klein, Eva Petkova, and Patrick Shrout have made to this research.

Correspondence concerning this article should be addressed to Colleen R. O'Neal or Laurie Miller Brotman, NYU Child Study Center, NYU School of Medicine, 215 Lexington Avenue, 14th Floor, New York, NY 10016. Electronic mail may be sent to onealc01@nyumc.org or millel02@nyumc.org.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8101-0019

parenting behavior, which, in turn, influences maternal–child interactions. This Gene × Environment effect could subsequently impact child stress response, which, in turn, may influence child behavior. From this perspective, relations among parent cortisol, parent behavior, child cortisol, and child behavior could each reflect the influence of both genes and the environment shared by parents and their offspring.

Experimental approaches provide unique opportunities to examine relations among family environment, child physiology, and behavior. The current study takes advantage of a randomized controlled prevention trial of an early family intervention among children at high risk for antisocial behavior. Previously, Brotman and colleagues demonstrated that the intervention impacts three aspects of parenting (responsive parenting, harsh parenting, and stimulation for learning), child engaging behaviors with unfamiliar peers, and child aggression (Brotman et al., 2005; Brotman et al., 2008), and that intervention effects on parenting are meaningfully related to the prevention of aggression during parent–child interactions (Brotman et al., 2009). Furthermore, the behavioral family intervention resulted in changes in children's cortisol response in anticipation of a social challenge with unfamiliar peers (Brotman et al., 2007), demonstrating the malleability of the stress response system among children at high risk for antisocial behavior. These reports did not examine how the family environment relates to child stress response or the relations between child cortisol and aggressive behavior. The primary aim of this study is to examine if intervention effects on child cortisol are meaningfully related to later child aggressive behavior, observed during a parent–child interaction 16 months after the intervention. To explore the specificity of this relation, the study also examines intervention effects on child engaging behavior during the same parent–child interaction and whether changes in cortisol mediate any observed intervention effects.

Cortisol Response and Behavior

Stress regulation is implicated in the development of psychopathology (e.g., Alink et al., 2008). Cortisol response to a stressor has been identified as an important indicator of stress regulation, with consequences for child behavior. However, as noted above, considerable literature demonstrates the complex nature of associations among cortisol, context, and behavior. Although low cortisol reactivity to a stressor (i.e., hypocortisolism) has been linked to aggressive behavior (Gunnar & Vazquez, 2001), careful analysis demonstrates more complex associations. For example, unique subgroups of children may exhibit positive associations between cortisol responding and aggression (e.g., Dadds & Rhodes, 2008; Hart et al., 2005; Rutter, 2008). Cortisol does not exhibit a simple one-to-one correspondence with aggression but, rather, can show positive associations in some contexts and negative associations in others. Theories explaining the process in which cortisol is linked to aggression differ. The fearlessness theory (Raine, 1993) posits that low levels of cortisol arousal or reactivity lead children to be fearless, and such fearless children are more likely to initiate fights because they are not concerned about consequences. Stimulation seeking theory (Zuckerman, 1979) argues that some children experience a low arousal, negative state, which leads them to seek higher levels of arousal via aggressive behavior. A third theory proposes that insensitivity to social and emotional distress signals (e.g., fear) in peers predisposes children to certain types of aggression (Blair, 2003).

Relative to typically developing children, children at high risk for psychopathology (e.g., children with familial antisocial behavior, maltreated children, or those raised in orphanages) or those with clinical diagnoses of disruptive psychopathology exhibit low or flattened cortisol responses to challenging situations as well as to circadian changes (Fisher, Gunnar, Chamberlain, & Reid, 2000; Gunnar & Vazquez, 2001; Hart, Gunnar, & Cicchetti, 1995). Considerable empirical literature relates low cortisol reactivity to aggression and disruptive psychopathology (Snoek et al., 2004; van Goozen et al., 1998; van Goozen et al., 2000). This literature, summarized in a recent meta-analysis, finds a medium effect size ($d = .57$) of low cortisol stress response for children characterized as antisocial or with clinical diagnoses of conduct disorder or oppositional defiant disorder (van Goozen, Fairchild, Snoek, & Harold, 2007). Although preschoolers were not included in the meta-analytic study, a study by Blair et al. (2005) found a similar relation between highly aggressive behavior and blunted cortisol reactivity among preschoolers from low-income families.

Despite the findings from one meta-analysis (van Goozen et al., 2007), some empirical evidence suggests either no significant relation between cortisol reactivity and aggression (see meta-analysis by Alink et al., 2008) or even a positive relation of cortisol reactivity to aggression (e.g., Hart et al., 2005). This lack of a simple one-to-one correspondence between stress response and aggression may be due to multiple complicating factors, including child age and type of disruptive psychopathology (Alink et al., 2008; Dadds & Rhodes, 2008). For example, children who exhibit physical aggression that is predatory or instrumental may be more likely to have blunted cortisol responses to stress than children with reactive aggression or those with more general nonaggressive disruptive psychopathology (Dadds & Rhodes, 2008; Loney, Butler, Lima, Counts, & Eckel, 2006; McBurnett, King, & Scarpa, 2003). The current study examines observed child behavior (physical aggression and socially engaging behavior) during a standardized parent–child interaction paradigm and cortisol levels in response to a social challenge among a relatively homogenous age group of children (33–63 months) followed prospectively over a 24-month period.

Experimental Studies of Child Neurobiology and Behavior

Because the literature on the stress response and aggressive behavior in humans is largely based on cross-sectional studies, there have been few advances in understanding the causal relations between cortisol response and child aggression. As a result, a call has been made for greater use of experimental studies that can illuminate causal mechanisms underlying the link between neurobiology and behavior (Rutter, 2008; Tremblay, 2008). Experimental studies of interventions that alter both the stress response and child behavior provide unique opportunities to examine the nature of the relation between cortisol response and behavior. In theory, restoring a high-risk child's physiological stress reaction to a typical, reactive state may lead to improved emotional and cognitive appraisal of a socially stressful situation, thus facilitating a less aggressive reaction and a potentially more positive, socially engaged reaction to stressful events (van Goozen et al., 2007).

Two experimental studies with maltreated children report changes in cortisol and child behavior (Dozier, Peloso, Lewis, Laurenceau, & Levine, 2008; Fisher, Stoolmiller, Gunnar, & Burraston, 2007). These studies considered diurnal patterns of cortisol secretion during typical days and did not examine the link between cortisol levels in response to stress and child behavior. A third study of older children (aged 8–13) with behavior disorders found that pretreatment cortisol response to stress moderated intervention outcomes on conduct problems (van de Wiel, van Goozen, Matthys, Snoek, & van Engeland, 2004). Since this study did not examine intervention effects on cortisol, the consequences of child stress response on child behavior remains unknown. In a prior report of immediate intervention outcomes on cortisol response in the current study sample, Brotman and colleagues found that intervention effects on cortisol levels in anticipation of a social challenge and engaged behavior with peers during the challenge were not related (Brotman et al., 2007). The current study extends this work and other prior experimental studies in this area by examining the links between immediate intervention effects on child cortisol response and longer term intervention effects on child engaging and aggressive behavior with the parent, with consideration of the family environment.

Early Family Environment and Stress Neurobiology

Stress-sensitive neurobiology is shaped by early experience, especially by caregiving (Gunnar & Quevedo, 2007). Family environmental factors that operate at key points in development might sculpt hypothalamic–pituitary–adrenal (HPA) axis regulation in children, much as environmental factors have been shown to play a role in HPA regulation in rodents and nonhuman primates (e.g., Caldji et al., 2000; Levine, 1957; Liu et al., 1997). Animal studies suggest both genetic and environmental factors involved in the transmission of abnormal stress regulation. Offspring at high risk for abnormal neuroendocrine stress regulation are characterized by having lower warmth parents (Meaney, 2001). Studies in young human children find evidence of parallel processes (Fisher, Gunnar, Dozier, Bruce, & Pears, 2006; Gunnar et al., 2006; Hane & Fox, 2006). Children from high-risk, stressful environments exhibit lower or flat diurnal cortisol levels compared to nonstressed children ("hypocortisolism"; Gunnar & Vazquez, 2001; Hart et al., 1995; Heim, Ehlert, & Hellhammer, 2000; Hertsgaard, Gunnar, Erickson, & Nachmias, 1995; Nachmias, Gunnar, Mangelsdorf, Parritz, & Buss, 1996). Although both positive and negative relations between quality of caregiving and child cortisol response to stress have been found among low-risk families (Blair & Diamond, 2008; Gunnar, Brodersen, Nachmias, Buss, & Rigatuso, 1996; Gunnar & Vazquez, 2001; Hane & Fox, 2006; Nachmias et al., 1996), studies of high-risk families consistently link suboptimal caregiving with low child cortisol response. For high-risk families, responsive parenting may serve as an important moderating risk or buffering factor for the development of hypocortisolism (i.e., low cortisol levels; Fisher et al., 2006; Gunnar & Vazquez, 2001).

In addition to the responsive nature of the caregiving environment, a small literature in humans suggests that parent stress responding and depression may be related to child cortisol response to stress in early childhood. A large twin study found modest similarity in afternoon basal cortisol levels between parents and children (Schreiber et al., 2006). A small study of diurnal levels of cortisol in typically developing children and their parents documented a positive correlation between diurnal levels of parent and child cortisol (Young, Vazquez, Jiang, & Pfeffer, 2006). Another study of low-risk families found a positive correlation between infant and mother diurnal cortisol, but the correlation was not significant between child and mother cortisol reactivity during a separation paradigm (Thompson & Trevathan, 2008).

Related to parent psychopathology, a small literature suggests that parent depression during early childhood may be related to child cortisol patterns (e.g., Lupien, King, Meaney, & McEwen, 2000). Parent depression in the first 2 years of life was found to predict child cortisol reactivity to a stressor at age 7 (Ashman, Dawson, Panagiotides, Yamada, & Wilkinson, 2002). In sum, parental warmth, parent stress response, and parent depression are likely to influence child cortisol response in early childhood. Yet, the aforementioned studies have not systematically characterized the family environment or examined parent and child factors among children from high-risk stressful environments. In the current study, therefore, we consider three aspects of the family environment (i.e., parental warmth, stress response, and depression) in relation to children's cortisol and behavior in a high-risk sample.

The Current Study

The current study takes advantage of a randomized controlled trial of an early family intervention among preschoolers at familial risk for antisocial behavior. In a series of studies, the intervention has been found to promote responsive parenting and child engaging behavior with peers, reduce physical aggression during parent–child interactions, and alter cortisol response in anticipation of a social challenge with peers (Brotman et al., 2005; Brotman et al., 2007; Brotman et al., 2008). Guided by current models of developmental psychopathology, we evaluate whether increased cortisol response is related to intervention effects on aggressive behavior with parents. To yield a more complete picture and to evaluate the specificity of the cortisol–aggression connection, we will also evaluate relations among cortisol response and engaging behavior with parents.

Method

Participants

This study includes children and parents who participated in a randomized controlled trial of a family intervention aimed at preventing early conduct problems and later antisocial behavior. The trial methodology and characteristics of participants are detailed elsewhere (Brotman, Gouley, O'Neal, & Klein, 2004; Brotman et al., 2005; Brotman et al., 2008). Briefly, all family court records of youths under the age of 16 adjudicated in two boroughs in New York City were screened between 1997 and 2001 for the presence of preschool-age siblings. "Sibling" was defined as a child who was biologically related to, or lived with, the adjudicated youth for at least 1 year, and whose caregiver also raised the adjudicated youth. Families were eligible if siblings were between 33 and 63 months, the caregiver spoke English, and the family lived within the designated geographic area. Families were ineligible if caregivers had a current substance use or psychotic disorder determined by structured diagnostic interview, or if preschoolers had pervasive developmental disorders or severe or profound mental retardation as reported by parents or indicated by standardized cognitive testing. Forty-eight percent of eligible families (*N* = 92) agreed to participate and completed assessments. Seven of the 92 families had more than one preschool-age child; none of the siblings were included in this study. Forty-seven families were randomized to the intervention condition and 45 families to the control condition that received no intervention.

Sixty-eight percent of preschoolers were biological siblings of the adjudicated youths (12% full; 56% half), 3% adoptive siblings, and 28% biological nephews, nieces, or cousins raised as "siblings." Eighty-three percent of caregivers (parents) were the preschoolers' biological mothers, 2% biological fathers, 10% grandmothers, 3% adoptive mothers, and 2% other female relatives. Parent mean age was 36.3 years (*SD* = 9.2). Forty-five percent of parents had not completed high school and 59% of families had household incomes under $15,000. Fifty-three percent of preschoolers were girls; 65% were African American, 27% Latino, and 8% other or mixed race. Child mean age was 47.52 months (*SD* = 8.28) and the average child IQ score (DAS; Elliott, 1990) was 83.2 (*SD* = 12.9). The stressors experienced by the sample are described elsewhere (Brotman et al., 2004); in addition to socioeconomic disadvantage and sibling antisocial behavior, children were exposed to high rates of maternal psychopathology (based on semistructured diagnostic interviews) and stressful life events (based on maternal reports).

Preventive Intervention

The prevention program was adapted from the Incredible Years Series (IYS), an efficacious program developed and tested extensively by Webster-Stratton (e.g., Webster-Stratton & Hancock, 1998). The program aimed to improve parenting practices and preschoolers' social competence with the goal of preventing later conduct problems. The program consisted of 6–8 months of active intervention (November–June), followed by a 3-month booster intervention (October–December) initiated approximately 4–6 months after the end of the intervention. The active intervention included 22 group sessions for parents, 22 group sessions for preschoolers, 22 guided parent–child interactions, and 10 home visits.

As per published IYS manuals, group leaders coached parents to use specific parenting skills (e.g., labeled *praise, time out*) and modeled, directed, and reinforced appropriate behavioral strategies. The program specifically encourages parents to use nonharsh, consistent, and appropriate disciplinary strategies, be less critical, and use positive reinforcement and play interactions to promote children's social competence. Groups for preschoolers incorporated elements from the Incredible Years Dinosaur Social Skills and Problem-Solving Curriculum (Webster-Stratton, 1990), which has been shown to increase the efficacy of the Parent Program (Webster-Stratton, Reid, & Hammond, 2004). Child group leaders taught social skills, reinforced positive behaviors, and provided consequences for negative behaviors. Home visits were designed to help parents implement skills at home. Leaders from the parent and children's groups together provided home visits following a curriculum corresponding to parent and preschool group lessons. Home visits created opportunities for developing and implementing behavior plans (e.g., placement of a time out chair in a crowded apartment).

The control condition consisted of assessments and monthly phone calls through which family service use and changes to family circumstances were evaluated. Parents were paid $5 for these monthly phone interviews. If families requested assistance with identifying social, educational, or mental health resources during these calls, a clinician followed up with information provided over the telephone.

Assessment Procedures

Participants were assessed four times, 8 months apart, over 24 months: baseline or preintervention (Time 1, Month 0), postintervention (Time 2, Month 8), 8-month follow-up (Time 3, Month 16), and 16-month follow-up (Time 4, Month 24; see Table 1). Previously, we reported main effects of the intervention on child cortisol response immediately postintervention (Brotman et al., 2007), and intervention effects on child aggression during a parent–child interaction at Time 4 (Brotman et al., 2008). The intervention took place between Time 1 and Time 2 and the booster sessions took place between Time 2 and Time 3; no intervention took place between Time 3 and Time 4. The length of time between assessments was not different for intervention and control conditions, $t(77) = -0.42$, $p > .10$. The preintervention assessment required three or four visits, and the three postintervention assessments required two visits each; one visit was at the study center and one was in the home.

Table 1
Timeline of Assessments

Time 1	Time 2	Time 4
0 months	8 months	24 months
4 years old	4 years 8 months old	6 years old
Preintervention (baseline)	Immediate postintervention	16 months postintervention
Child cortisol reactivity	Child cortisol reactivity	Child aggression
Parent cortisol reactivity		Child engaging behavior
Parent depression		
Parent warmth		

Measures

Child cortisol in anticipation of a social challenge. Neuroendocrine functioning was measured by collecting salivary cortisol samples prior to an unfamiliar peer situation (prechallenge cortisol). Cortisol assessments were collected at baseline prior to randomization (Time 1) and repeated in an identical fashion immediately after intervention (Time 2). Samples were collected using a sterile cotton roll held by dental floss. The child mouthed the roll until it was wet. The sample was then placed in a vial and kept in a freezer until transported on ice to the laboratory. Food intake was controlled during the collection process.

The social challenge involved entry into an unfamiliar peer group for 30 min of play at a nursery school several blocks from the study center. Observations were scheduled with the school to correspond with planned free play. Approximately, 40 children (ages 34–60 months) from diverse socioeconomic and ethnic backgrounds were enrolled in the school. During free play, children were free to move from station to station (e.g., blocks, pretend house, art table) in a large room designed for this purpose. At least one teacher and three assistants from the school staffed free play.

Prior to leaving the study office for the school, a research assistant unaware of intervention status and study hypotheses explained to the child (in front of the parent) that he or she was "going to a school to play with some other children." Thus, children were informed about an impending potentially stressful experience prior to cortisol sampling. Two or three research assistants escorted the child and parent to the school (a 10– to 15–min walk). On arrival, and in the presence of the parent, a salivary cortisol sample was collected. The sample was collected approximately 15–20 min after the child was told about the challenge task. Because cortisol release in preschoolers occurs as early as 10 min from the stimulus presentation, we consider this cortisol sample to be an index of the child's *anticipation* of the challenge. After the cortisol sample was taken at the school, the child was then required to separate from the parent and accompanied to the play area by one or two research assistants who observed the child during 30 min of free play. Assessment procedures, including timing of cortisol collection, were identical for intervention and control conditions and for both assessment points. For Times 1 and 2, children's cortisol samples were collected between 10:00 and 11:30 a.m. or between 2:00 and 4:00 p.m., with the majority collected in the afternoon. Although individual children were not assessed at the same exact time at both assessments, major differences in levels based on diurnal patterns are not expected in this time range from late morning to early afternoon. Further details of the observational paradigm and intervention effects on behavior observed during this procedure are presented elsewhere (Brotman, Gouley, & Chesir-Teran, 2005; Brotman et al., 2005).

Children's cortisol levels in anticipation of the social challenge may reflect the child's response to a number of cues including separation from the parent, entry into a social peer situation, or a general transition. To further understand the challenge data, we also collected cortisol samples at home over the course of the day to evaluate diurnal patterns. We examined intervention effects on diurnal patterns but did not find an effect. Details on the diurnal patterns and the relation between cortisol in anticipation of a social challenge and cortisol levels over the course of the day at home are reported in Brotman et al. (2007).

Observed child behavior during parent–child interaction. Child behaviors with the parent were observed by raters unaware of intervention status during semistructured, parent–child interactions at the home and study center at Times 1–4. For this study, we utilize data from Times 1 and 4 only. At each assessment point, two parent–child interaction measures were obtained, a 10-min session at the family's home (5-min child-directed free play and 5-min parent-directed puzzle task), and a 15-min session at the study center (7-min child-directed free play, 5-min parent-directed puzzle task, and 3-min cleanup). Home and center-based assessments took place approximately 1 week apart. Home assessments were coded live and assessments in the study center were videotaped and coded at a later time. Different raters coded home and center-based observations and were uninvolved in other aspects of the study.

The Dyadic Parent–Child Interaction Coding System–Revised (DPICS–R; Robinson & Eyberg, 1981; Webster-Stratton, 1985) was used to record the frequency of discrete aggressive behaviors in both settings. Observed child physical aggression was derived from the two DPICS–R codes of physical aggression: child physical negative (i.e., child is physically aggressive toward parent) and child destructive (i.e., child destroys, damages, or attempts to damage objects). The two codes of physical aggression were significantly related ($r = .43$, $p < .001$) and combined. Reliability estimates consider average reliability across assessment time points. To estimate interrater reliability, approximately 20% of observations (in each setting, at each time point, and in each condition) were coded by two raters. Interrater agreement, calculated via single-measure intraclass correlation coefficients with a one-way random-effects model (Shrout & Fleiss, 1979), was .85 for direct home observations and .84 for videotaped center observations. Mean level aggression scores across home and center settings were not significantly different and were highly correlated ($r = .67$, $p < .001$). At each time point, we created a composite index of physical aggression by combining ratings across the two settings. The total physical aggression score reflects the number of times the child aggressed toward his or her parent or destroyed objects over a total of 25 min. The estimated reliability of the composite across all time points was .88 (Lord & Novick, 1968).

Although the primary focus of this study is on aggressive behavior, we also examined engaging behavior with the parent to gain a more complete picture of behavioral outcomes. Of note, although we have reported immediate intervention effects on engaging behavior with peers (Brotman, Gouley, Chesir-Teran, & Dennis, 2005), we have not previously examined longer term intervention effects (Time 4) on engaging behavior with the parent. The Global Impressions of Parent–Child Interactions (GIPCI; Brotman, 1998) was used to make global ratings of child positive, engaging behavior with the parent. GIPCI ratings were made on a 5-point scale using descriptions of prototypic behaviors. Child positive engagement was the average of the following three GIPCI scales: child emotion (i.e., positive emotional tone of the child behaviors), child play with the parent (i.e., child shares and encourages activity with parent, shows appropriate verbal skills and cooperation), and reverse-coded child disconnection (i.e., child is withdrawn, aimless, requires encouragement to play and relate to parent). The three codes of positive engagement were significantly related ($rs = .53–.69$) and were combined. We combined home and study center observations because both were related ($r = .38$ at Time 1 and $r = .44$ at Time 4). Test–retest reliability for the combined scale ranged from .47 to .61. The ICC for the combined composite is .75. The estimated reliability of the composite was .83 (Lord & Novick, 1968).

Parent warmth. The Home Observation for the Measurement of the Environment–Early Childhood Version (HOME–EC; Caldwell & Bradley, 1984) was used to assess aspects of the home environment and parenting. The observation of the parent–child relationship occurred in the home over the course of an approximately 2-hr visit. For the current study, we used warmth data at Time 1. Items are rated by interviewers as present (1) or absent (0). We employed the conventional Caldwell and Bradley HOME Warmth scale (Caldwell & Bradley, 1984). It is a seven-item scale including the following items: "Parent holds child close 10–15 min per day"; "Parent converses with child at least twice during visit"; "Parent answers child's questions or requests verbally"; "Parent usually responds verbally to child's speech"; "Parent praises child's qualities twice during visit"; "Parent caresses, kisses, or cuddles child during visit"; and "Parent helps child demonstrate some achievement during visit." The summary score was the mean of all seven items. Internal consistency was moderate ($\alpha = .61$). Approximately one third of the sample had scores of 1 on all seven items (i.e., all seven items were observed in the home). For the remainder of the sample (67%), scores ranged from 0 to 6.

Parent depression. We assessed parent depression using the Structured Clinical Interview for DSM–IV nonpatient version (SCID–NP; Williams et al., 1992) by trained clinical psychologists. An independent clinician reviewed all clinical interviews for diagnostic decision making by interviewers. Psychometric properties of the SCID are adequate in other samples (Williams et al., 1992).

Parent cortisol in anticipation of a challenge. Parent neuroendocrine functioning was measured by collecting salivary cortisol samples prior to the administration of a Wechsler Adult Intelligence Scale–Vocabulary subtest and a Wide Range Achievement Test–Reading subtest (i.e., prechallenge cortisol). Cortisol assessments were collected prior to randomization (Time 1) on the same day and within a 1-hr time period of when the child provided his or her sample. Parents were informed that they were going to be asked to complete a test of their vocabulary and reading skills. Saliva samples were taken approximately 10 min after the instructions were given; as such, we consider this cortisol sample to be an index of the parent's *anticipation* of the challenge. The literature supports that cognitive testing should serve as a stressful challenge for adults in a similar fashion as the peer entry paradigm is stressful for young children. A cognitive test may be perceived as especially challenging for parents in this sample since their mean level of cognitive functioning and achievement was well below average (Brotman et al., 2004). Samples were collected using a sterile cotton roll held by dental floss. The parent mouthed the roll until it was wet. The sample was then placed in a vial and kept in a freezer until transported on ice to the laboratory. Food intake was controlled during the collection process. The assessment of parent cortisol was designed so that it would yield information that would be helpful in interpreting child cortisol levels. As with child cortisol, parent cortisol was also collected from parents at home at four time points over the course of the day (the same day and time points as the child). We do not report these data here, but we evaluated these diurnal levels to aid interpretation of challenge data. Diurnal levels are available from the authors.

Analyses

Attrition and imputation strategies. Time 1, 2, and 4 data were utilized for the current study. There was modest attrition by Time 2 (87% retained) and Time 4 (77% retained). There were no differences by intervention status in the percentage of children and parents missing data at any time. Furthermore, there were no baseline differences between those with complete data versus those with missing data at later times, and there were no significant baseline score by intervention status interactions in predicting missingness. Given these results, we assume that the data are missing at random (Schafer & Graham, 2002).

As per Little and Rubin (2002), we used the SAS Multiple Imputation (SAS Institute, Cary, NC) procedure with 10 replicated imputations to impute data for all study variables; the imputation made use of the joint distribution of all available measures of parent and child functioning. Imputations were conducted separately for the control and intervention groups to account for the possibility of different missing data patterns by group (Little & Rubin, 2002). SAS PROC MIANALYZE was used to combine the results of the 10 imputations for the final inference. This combination of multiple imputations provides estimates of standard errors that take into account the fact that some data were imputed. Of note, we did not impute missing data for observed child physical aggression because the distribution of scores was highly skewed, even after log transformation. The imputation methods are optimal for normally distributed variables and are not appropriate for highly skewed variables. Therefore, we used the approach of carrying the last time point forward to replace the missing values for this variable. All analyses yielded highly similar results regardless of using the imputed or nonimputed data set.

Analytic approach. Correlations were estimated and tested to determine the concurrent relations among aspects of the family environment, child cortisol response, and child behavior. We tested a mediation process in which child cortisol response was expected to mediate the intervention effect on later child outcomes (i.e., aggression, engaging behavior): The intervention effect on child cortisol response was expected to be related to the intervention effect on child outcomes. We employed the "joint significance test" strategy in which mediation is inferred if there is evidence that both paths *a* (intervention effect on cortisol) and *b* (change in cortisol and later behavior) in Figure 1 are significant (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002). We also carried out the conventional Sobel test described by Baron and Kenny to test the significance of the mediation model (1986).

To test the percent of the intervention effect accounted for by the mediator, we used an equation that is appropriate for a nonlinear mediation model. In a linear-based mediation

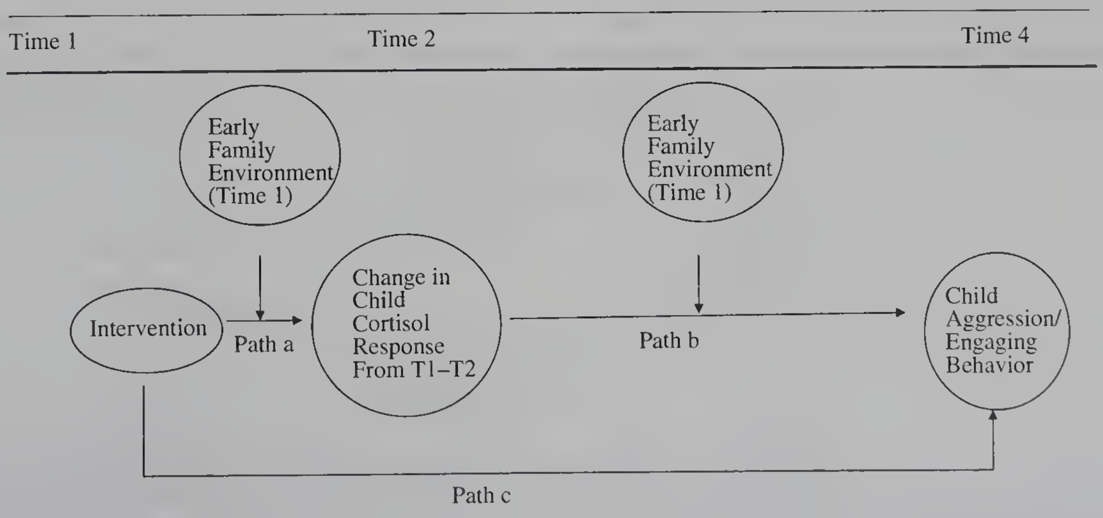

Time 1
Time 2
Time 4
Early Family Environment (Time 1)
Early Family Environment (Time 1)
Intervention
Path a
Change in Child Cortisol Response From T1–T2
Path b
Child Aggression/ Engaging Behavior
Path c

Figure 1. Hypothesized mediation model.

model, the total intervention effect (c) is equal to the sum of the part that is explained by the mediator (paths a and b) and the part that does not involve the mediator (c'). In a nonlinear mediation model, like our model, c is not equal to the total intervention effect (i.e., c is not equal to the sum of the part explained by the mediator [a and b] and the part that does not explain the mediator [c']; Fleiss, Levin, & Paik, 2003). For this reason, we compute the total effect as the indirect effect/total effect $= (a \times b)/[(a \times b) + c']$.

Poisson regression analysis, a log-linear approach appropriate for low frequency behavior counts with skewed distributions (Fleiss et al., 2003), was used to test the relation between cortisol change and aggression. The Poisson regression model was carried out using a generalized estimating equation approach (Diggle, Heagerty, Liang, & Zeger, 2002; Liang & Zeger, 1986; Zeger, Liang, & Albert, 1988) with a log-link function (Diggle et al., 2002). PROC GENMOD in SAS® (SAS Institute, Cary, NC.) was used to perform the analyses.

A mediation model is ideally tested in a design demonstrating time precedence in which the intervention occurs before the mediator and the mediator occurs before the outcome (Cole & Maxwell, 2003; Kraemer, Wilson, Fairburn, & Agras, 2002; Sobel, 1990). Note that we only examined the outcomes at Time 4, rather than modeling change over multiple time points, due to the temporal limitations of testing a mediation model.

Upon inspection of the distribution of child aggression scores across the range of parental warmth, we found that there was essentially no child aggression at Time 1 or Time 4 among the families with scores of 1 on warmth (33% of the sample). Among the high-warmth families (warmth = 1), only 1 child in the control group and 1 in the intervention group had nonzero scores at baseline. At Time 4, 2 children in the control condition had nonzero aggression scores and all children in the intervention had 0 aggression scores. As a result, we eliminated the families with scores of 1 on warmth from the mediation analyses relating child cortisol to aggression.

Results

Preliminary Analyses

Descriptive means and standard deviations for the sample by intervention are presented in Tables 2 and 3. Because both tables are intended to provide information necessary for interpretation of the mediation analyses on aggression, all descriptive information is presented with the removal of the families with scores of 1 on warmth at baseline. As shown in Table 2, no baseline (Time 1) intervention group differences were found for any of the child or family variables.

Relations Among the Family Environment, Cortisol Response, and Child Behavior

Table 3 shows concurrent, baseline relations among aspects of the family environment (parent cortisol response, warmth, and depression), child cortisol response, and child behavior.

Intervention Effect on Cortisol (Path a)

Previously, Brotman et al. (2007) reported intervention effects on cortisol for the entire sample. Since we removed a subset of families with high

Table 2
Descriptives (Mean and Standard Deviation) by Intervention Status

	Control (N = 24)		Intervention (N = 33)	
	Time 1	Time 2	Time 1	Time 2
Child cortisol	0.07 (0.31)	−0.09 (0.33)	0.07 (0.32)	0.27 (0.50)
Parent cortisol	0.11 (0.27)		0.05 (0.29)	
Parent warmth	0.66 (0.18)		0.67 (0.16)	
Parent depression	42%		35%	
	Time 1	Time 4	Time 1	Time 4
Child aggressive behavior	0.24 (0.79)	1.05 (3.73)	0.86 (2.41)	0.26 (0.71)
Child engaging behavior	3.96 (0.50)	4.02 (0.56)	3.81 (0.38)	3.91 (0.44)

Note. Table does not include families with warmth = 1.

Table 3

Descriptives and Correlations Among Baseline Early Family Environment, Child Cortisol Response, and Child Behavior

	N	*M* (*SD*)	Child cortisol	Parent cortisol	Parent warmth	Parent depression	Child aggression
Child cortisol	44	0.07 (0.31)					
Parent cortisol	47	0.08 (0.28)	.25				
Parent warmth	57	0.66 (0.17)	.08	.21			
Parent depression	56	0.39 (0.49)	−.08	.01	−.14		
Child aggression	57	0.50 (1.69)	−.17	−.07	−.21	−.04	
Child engaging	57	3.90 (0.45)	−.22	.21	.24†	.25†	−.24†

Note. N for correlations in the low warmth group ranged from 40 to 57. Table does not include families with warmth = 1.
$^{\dagger}p < .10$.

warmth scores from the mediation analyses, we first needed to replicate the intervention effect on cortisol among the remaining families. Results indicate that the intervention had a significant effect on cortisol, Estimate (*SE*) = .49 (.20), $p < .05$. Similar to findings previously reported for the full sample, cortisol levels increased in the intervention condition (.07 to .27) and decreased in the control group (.07 to −.09; see Table 2). Even after removing the high-warmth families from these analyses, maternal warmth at baseline significantly moderated the intervention effect on cortisol, Estimate Intervention × Warmth (*SE*) = 1.83 (.61), $p < .01$. The pattern of findings shows that the intervention had a stronger effect on cortisol among the families with lower warmth scores at baseline.

Previously, we reported long-term intervention effects on parent responsiveness for the entire sample (Brotman et al., 2008). For this study, we reexamined the intervention effect on parental warmth within the smaller sample (without the highest warmth families and with a slightly different measure so that it was consistent with the baseline measure of warmth considered in this study) and then conducted a post hoc analysis of the relation between intervention effects on parental warmth and child cortisol. As expected, the intervention significantly increased warmth ($\beta = .15$, *SE* = .06, $p = .01$), and there was a positive, although nonsignificant, association between change in warmth and change in cortisol ($r = .24$, $p = .21$), such that increases in warmth were associated with increases in cortisol response.

Cortisol Change and Later Aggression (Path b)

There was a significant relation of change in cortisol from pre- to postintervention with child aggression at Time 4, Estimate (*SE*) = −1.96 (.62), $p = .007$ (see Figure 2). The relation between cortisol change and later aggression was not moderated by warmth, Estimate intervention × warmth (*SE*) = 1.88 (1.35), $p = .17$.

Taken together, these results meet the joint significance test for mediation of child aggression because both paths *a* (intervention effects on cortisol) and *b* (change in cortisol and later behavior) were significant (MacKinnon et al., 2002).

Intervention Effect on Child Aggression (Path c)

Brotman and colleagues previously reported that there was a significant intervention effect on Time 4 aggression across the entire sample (Brotman et al., 2008). After the removal of the high warmth families,

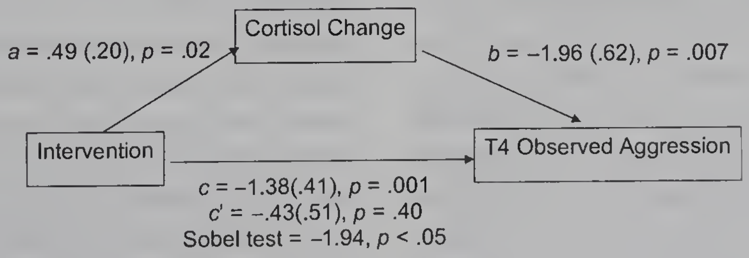


Figure 2. Estimates for indirect intervention effect for lower warmth group (warmth < 1).
Note. N = 57; subsample of high warmth (warmth score = 1) has been removed. a = intervention effect on cortisol change; *b* = cortisol change on aggression; *c* = total intervention effect on aggression; *c′* = total intervention effect on aggression, adjusting for the mediator (cortisol change).

there was a significant intervention effect on Time 4 aggression, $c = -1.38$ (.41), $p = .001$ (see Figure 2). This effect was not moderated by warmth, Estimate Intervention × Warmth (*SE*) = 8.03 (5.91), $p = .20$.

Intervention and Child Aggression Adjusted for Mediator (Path c′)

The results from the previous three sections show that the mediation link between the intervention, cortisol, and later aggression was supported (paths *a* and *b* were significant; see Figure 1). In this section, we test paths *c* and *c′* (whether the effect of intervention on aggression is no longer significant after adjusting for the mediator). As described above, the intervention effect on aggression (path *c*) was significant, $c = -1.38$ (.41), $p = .001$. As expected in complete mediation, regression analyses showed that after adjusting for the mediator (change in cortisol), the intervention was no longer significantly correlated with later aggression, $c' = -.43$ (.51), $p = .40$.

Post hoc analyses revealed that changes in cortisol accounted for 69% of the intervention effect on child aggression. We find further support for the mediating role of cortisol on aggression using the Sobel test (Sobel test = −1.94, $p < .05$; see Figure 2).

To summarize, after removing the highest warmth families because their children did not display aggression at baseline or Time 4, there was an intervention effect on cortisol (that was strongest for the lower warmth families) and intervention-induced change in cortisol response mediated the intervention effect on aggression. Over two thirds of the intervention effect on Time 4 aggression during a parent–child interaction was accounted for by the intervention effect on cortisol response in anticipation of a social challenge.

Child Engaging Behavior

Across the entire sample, we documented a significant intervention effect on later child engaging behavior with the parent, 16 months after the intervention, Estimate (*SE*) = .26 (.08), $p = .003$. To be able to compare results from the mediation analyses with those for child aggression, we limited the analyses of engaging behavior to the lower warmth sample only. The intervention effect on engaging behavior was not significant among lower warmth families, Estimate (*SE*) = .12 (.15), $p = .41$, and a change in cortisol from pre- to postintervention was not related to child engaging behavior at T4, Estimate (*SE*) = .09 (.42), $p = .33$; since path *b* was not significant, mediation did not occur for engaging behavior via cortisol change.

Discussion

The current study takes advantage of an experimental trial with young children at high risk for antisocial behavior to further understand the complex nature of relations among the family environment, neurobiological regulation of stress, and developmental psychopathology. In earlier work with this high-risk sample, we found that family intervention that successfully prevented aggression during early childhood (Brotman et al., 2008) also resulted in normalized stress response (Brotman et al., 2007). The current study extends these findings by examining associations between immediate intervention effects on cortisol response and intervention effects on later aggression, with consideration of the family environment. Study findings underscore the importance of parental warmth for the development of cortisol response during early childhood, suggest that the cortisol response system is plastic (amenable to change as a result of psychosocial intervention) in children from high-risk families and those further characterized by lower maternal warmth, and indicate that cortisol response in anticipation of a social stressor predicts later child aggression. Among the subset of families considered in this study of meditational processes, the long-term prevention effect on child aggression was largely explained by immediate intervention effects on cortisol levels. These findings hold important theoretical and practical implications for the development of the stress response, developmental psychopathology, and its prevention.

Although low cortisol has been associated with disruptive psychopathology, studies demonstrating this link remain largely correlational in nature and, for the most part, have been conducted with children already exhibiting problems (Alink et al., 2008). Furthermore, studies linking child cortisol and behavior have not typically characterized the family environment in which these associations are found. The current study investigates the connection between child cortisol and aggression, while also considering family environment. It links intervention-related increases in cortisol in anticipation of a social challenge to the prevention of physical aggression among children at high risk for maladjustment. Findings are consistent with the theoretical proposition that prevention of aggressive behavior, among children at high risk for atypical stress responding, is dependent on changing the stress response system (van Goozen et al., 2007).

The results of this study beg future in-depth probing of the underlying mechanisms linking changes in cortisol to reduced aggression. Of particular interest in this study is how children's anticipation of a social challenge with peers relates to later aggressive behavior during a parent–child play interaction. A theoretical model worth testing is whether or not child physiological stress response impacts cognitive and emotional processes leading to aggression (van Goozen et al., 2007). In theory, restoring a child's physiological stress response to a typical, reactive state may lead to improved emotional and cognitive appraisal of a socially stressful situation, and these skills generalize across settings and over time, resulting in less aggressive reactions (van Goozen et al., 2007). An alternative but related explanation of results may be that children's increased, more typical, reaction to stress may make them more sensitive to stress cues in relationships, which may, in turn, make them less aggressive toward others. Perhaps such increased sensitivity to social stress leads children to be more fearful of the negative consequences of misbehavior, resulting in lower rates of aggression (Raine, 1993). It has also been posited that heightened stress reactivity occurs due to increased "biological sensitivity to context" with negative consequences in an adverse context and positive consequences in a supportive context (Boyce & Ellis, 2005; Ellis & Boyce, 2008). An intervention could act as a supportive context for families. Perhaps, in this sample, a positive behavioral outcome (i.e., low levels of aggressive behavior) as a consequence of increased stress reactivity was especially possible due to the supportive context an intervention offers to a family.

The current study points to the importance of how the early family environment relates to child cortisol response to a stressor. This study's result linking parent and child cortisol response is too preliminary to speculate as to whether or not the association is due to genetic or learned causes; however, it will be important to identify mechanisms leading children to develop cortisol reactivity patterns that are similar to their parents. Future studies may want to examine whether the most direct route to understanding child neurobiology is not via parent psychopathology or parenting, but via parent neurobiology. In addition to the clustering of parent depression, warmth, cortisol, and child engaging behavior and associations between parent and child cortisol, preliminary, post hoc analyses suggested that among lower warmth families, where changes in both responsive parenting and cortisol levels were observed, responsive parenting and child cortisol response changed in a parallel fashion. Together, the sum picture of such correlational results suggests the importance of research testing causal pathways explaining the link between family environment and child neuroendocrine response during the important developmental period of early childhood.

Given that the link between neurobiological change and aggression occurred among lower warmth families in this study, the intervention effect on cortisol and aggression suggests that caregiving is an important element of a compensatory resilience process (see Luthar, Cicchetti, & Becker, 2000) promoting healthy early neurobiological growth among high-risk children. Results suggest that the early family environment of warmth is a significant contextual factor in explaining change in cortisol response and the consequence of such cortisol change on child aggression. Results contribute to the small literature demonstrating biological outcomes from early intervention with young children (Brotman et al., 2007; Dozier et al., 2008; Fisher et al., 2007; Raine et al., 2001). Our findings underscore the importance of considering characteristics of the family environment (e.g., parental warmth) within studies that examine biological and behavioral outcomes from family intervention (Cicchetti & Gunnar, 2008). In intervention children from lower warmth families, cortisol levels in anticipation of a social challenge were found to increase over an 8-month period, from approximately ages 4–5 years. In contrast, in control children from lower warmth families, cortisol levels decreased over this same developmental period. These findings support a developmental model in which moderately low levels of parent warmth place children who are at familial risk for antisocial behavior at even further risk for an atypical stress response. In these lower warmth families, preventive intervention that enhances parenting practices and child social competence (Brotman et al., 2005) may protect against this risk; immediately following intervention, children from lower warmth intervention families exhibited increases in cortisol relative to controls. It is important that future research more definitively establishes the developmental trajectory of normal cortisol response patterns over time among young children and the specific challenge conditions that elicit such a response. Translation of basic research to preventive interventions is dependent on future research, which details the stress buffering mechanisms in the family environment leading to positive neurobiological and behavioral outcomes in high-risk children (Gunnar et al., 2006).

This study also extends previous research (e.g., Fisher et al., 2006) by characterizing the family environment and the subgroups that benefit from intervention. Additional post hoc results suggested that among lower warmth families, the intervention altered both parent warmth and child cortisol, and these changes may be related. Further research is needed to more fully understand how change in parenting may lead to change in child neuroendocrine functioning and behavior. The unique importance of warmth for the development of the neurobiological system is strongly supported by animal studies (e.g., Caldji et al., 2000; Liu et al., 1997), especially an elegant experimental literature of rodent maternal care on rat pup stress response. Unlike the physical process underlying rodent parent warmth (i.e., licking and grooming of their offspring), the measure of human parental warmth as assessed in this study largely consisted of verbal behaviors. It is not known whether findings are specific to verbal behaviors or whether inclusion of physical behaviors (e.g., hugging, physical proximity) would provide a more complete picture of this process. As these lower warmth families were also characterized by parent depression and hypocortisolism, future exploration of the construct of warmth and its relation to other aspects of parenting behavior and neurobiology in high-risk families appears warranted.

Future studies should also consider additional outcomes of child functioning and parent–child relations. In the current study, increased cortisol in anticipation of a social challenge was related to lower rates of aggressive behavior during parent–child interactions. It is possible that higher cortisol levels may be related to other outcomes, including negative outcomes such as poor physical health. In this study, aggression, not child engaging behaviors, was related to child cortisol change. Perhaps other positive child outcomes, such as child emotion coping or academic success, would be affected by change in child cortisol response. Also, in empirical studies of cortisol and aggression, it has been argued that it is best to examine cortisol under stressful conditions (i.e., cortisol reactivity) rather than relying on diurnal cortisol (van Goozen et al., 2007); the cortisol patterns unique to aggressive behavior may be specific to or more easily detected in stressful or challenging conditions (McBurnett, Lahey, Rathouz, & Loeber, 2000; Pajer, Gardner, Rubin, Perel, & Neal, 2001; Popma et al., 2006; van Goozen et al., 2000). Therefore, future studies may want to determine what kinds of outcomes are uniquely affected by change in cortisol response to challenge versus diurnal patterns of cortisol. Similarly, the specific context of the cortisol stressor could predict different subtypes of aggression. For instance, if the cortisol response paradigm tested reactivity or lack of reactivity to others' distress, rather than reactivity to novel peers as in this study, then the lack of reactivity to others' distress might better predict emotional callousness (Dadds & Rhodes, 2008).

In addition to further understanding relations among family environment, child cortisol response, and behavior, this study provides further evidence in support of early family intervention for high-risk children. Specifically, new findings on child outcomes in this study include intervention effects on observed child engaging behavior, 16 months after intervention.

There are a number of important study limitations that are worthy of consideration. First, we had to eliminate families with the highest scores on warmth from the mediation analyses. As a consequence, none of the findings can be generalized to families characterized by high warmth. Future studies should conduct mediation and moderation analyses of intervention effects on aggression to formally test many of the proposed hypotheses. In addition, the generalizability of study findings is limited due to the unique nature of the sample. In terms of the sample characteristics, attempts to replicate the current findings should consider whether similar effects on cortisol can be documented among children living in environments involving different forms of stress. The current findings should encourage attempts to examine effects on physiology in much larger samples, where unique subgroups can be examined more comprehensively. Despite the need for replication, the current findings demonstrate statistically significant mediating effects. This suggests the potential value of conducting future work examining mechanisms linking context, behavior, and stress response in children. Such future work might include longer term outcomes assessed over multiple time points.

As with the initial report on intervention effects on cortisol response (Brotman et al., 2007), the current study is limited by the lack of a low-risk comparison group. Including such a group, studied with identical methods, is key given the robust effects of context on both behavior and cortisol. However, the association between increased cortisol response and lower rates of aggression suggest that intervention effects on cortisol response were positive and that the intervention worked to normalize the stress response. The study is also limited by its reliance on one measure of parent warmth. It would have been desirable to have a more comprehensive measure of this construct that includes both verbal and physical aspects of warmth. Also, we do not know whether study findings generalize to child behavior outside the home environment. That is, we did not measure behavior with peers, and, therefore, it is not known whether the relation between cortisol and behavior is limited to the context of parent–child interactions.

Much remains to be explained about the role of the early family environment, stress response, and the prevention of psychopathology in high-risk children. It is imperative that prevention studies continue to examine neurobiological as well as behavioral systems and outcomes (Cicchetti & Gunnar, 2008). Additionally, we would urge investigators to consider many aspects of the caregiving environment as potential moderators of these mediational processes (Fairchild & MacKinnon, 2009). Such steps would advance the field considerably in understanding the complex interplay of the early family environment, neurobiological functioning and the development, and prevention of aggression and psychopathology.

References

Alink, L., van IJzendoorn, M. H., Bakermans-Kranenburg, M. J., Mesman, J., Juffer, F., & Koot, H. M. (2008). Cortisol and externalizing behavior in children and adolescents: Mixed meta-analytic evidence for the inverse relation of basal cortisol and cortisol reactivity with externalizing behavior. *Developmental Psychobiology, 50,* 427–450.

Ashman, S. B., Dawson, G., Panagiotides, H., Yamada, E., & Wilkinson, C. W. (2002). Stress hormone levels of children of depressed mothers. *Development and Psychopathology, 14,* 333–349.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182.

Blair, R. J. (2003). Neurobiological basis of psychopathy. *British Journal of Psychiatry, 182,* 5–7.

Blair, C., & Diamond, A. (2008). Biological processes in prevention and intervention: The promotion of self-regulation as a means of preventing school failure. *Development and Psychopathology, 20,* 899–911.

Blair, C., Granger, D., & Razza, R. P. (2005). Cortisol reactivity is positively related to executive function in preschool children attending Head Start. *Child Development, 76,* 554–567.

Boyce, W. T., & Ellis, B. J. (2005). Biological sensitivity to context: I. An evolutionary-developmental theory of the origins and functions of stress reactivity. *Development and Psychopathology, 17,* 271–301.

Brotman, L. M. (1998). *GIPCI: Unpublished manual,* New York University School of Medicine.

Brotman, L. M., Gouley, K. K., & Chesir-Teran, D. (2005). Assessing peer entry and play in preschoolers at risk for maladjustment. *Journal of Clinical Child and Adolescent Psychology, 34,* 671–680.

Brotman, L. M., Gouley, K. K., Chesir-Teran, D., Dennis, T., Klein, R. G., & Shrout, P. (2005). Prevention for preschoolers at high risk for conduct problems: Immediate outcomes on parenting practices and child social competence. *Journal of Clinical Child & Adolescent Psychology, 34,* 724–734.

Brotman, L. M., Gouley, K. K., Huang, K.-Y., Kamboukos, D., Fratto, C., & Pine, D. S. (2007). Effects of a psychosocial family-based preventive intervention on cortisol response to a social challenge in preschoolers at high risk for antisocial behavior. *Archives of General Psychiatry, 64,* 1172–1179.

Brotman, L. M., Gouley, K. K., Huang, K.-Y., Rosenfelt, A., O'Neal, C. R., Klein, R. G., et al. (2008). Preventive intervention for preschoolers at high risk for antisocial behavior: Long-term effects on child physical aggression and parenting practices. *Journal of Child Clinical and Adolescent Psychology, 37,* 386–396.

Brotman, L. M., Gouley, K. K., O'Neal, C., & Klein, R. G. (2004). Preschool-aged siblings of adjudicated youths: Multiple risk factors for conduct problems. *Early Education & Development, 15,* 387–406.

Brotman, L. M., O'Neal, C. R., Huang, K.-Y., Gouley, K. K., Rosenfelt, A., & Shrout, P. (2009). An experimental test of parenting practices as a mediator or early childhood physical aggression. *Journal of Child Psychiatry and Psychology, 50,* 235–245.

Caldji, C., Diorio, J., & Meaney, M. J. (2000). Variations in maternal care in infancy regulate the development of stress reactivity. *Biological Psychiatry, 48,* 1164–1174.

Caldwell, B. M., & Bradley, R. H. (1984). *Home Observation for Measurement of the Environment–Revised Edition.* Little Rock: University of Arkansas at Little Rock.

Champagne, F. A., & Meaney, M. J. (2006). Stress during gestation alters postpartum maternal care and the development of the offspring in a rodent model. *Biological Psychiatry, 59,* 1227–1235.

Cicchetti, D., & Gunnar, M. R. (2008). Integrating biological measures into the design and evaluation of preventive interventions. *Development and Psychopathology, 20,* 737–743.

Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology, 112,* 558–577.

Dadds, M. R., & Rhodes, R. (2008). Aggression in young children with concurrent callous-unemotional traits: Can the neurosciences inform progress and innovation in treatment approaches? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*, 2567–2576.

Diggle, P., Heagerty, P., Liang, K.-Y., & Zeger, S. (2002). *Analyses of longitudinal data*. Oxford, UK: Oxford University Press.

Dozier, M., Peloso, E., Lewis, E., Laurenceau, J. P., & Levine, S. (2008). Effects of an attachment-based intervention on the cortisol production of infants and toddlers in foster care. *Development and Psychopathology*, *20*, 845–859.

Elliott, S. N. (1990). *Differential Abilities Scale*. San Antonio, TX: Psychological Corporation.

Ellis, B. J., & Boyce, W. T. (2008). Biological sensitivity to context. *Current Directions in Psychological Science*, *17*, 183–187.

Fairchild, A. J., & MacKinnon, D. P. (2009). A general model for testing mediation and moderation effects. *Prevention Science*, *10*, 87–99.

Fisher, P. A., Gunnar, M. R., Chamberlain, P., & Reid, J. B. (2000). Preventative intervention for maltreated preschoolers: Impact on children's behavior, neuroendocrine activity, and foster parent functioning. *Journal of the American Academy of Child and Adolescent Psychiatry*, *39*, 1356–1364.

Fisher, P. A., Gunnar, M. R., Dozier, M., Bruce, J., & Pears, K. C. (2006). Effects of therapeutic interventions for foster children on behavioral problems, caregiver attachment, and stress regulatory neural systems. *Annals of the New York Academy of Sciences*, *1094*, 215–225.

Fisher, P. A., Stoolmiller, M., Gunnar, M. R., & Burraston, B. O. (2007). Effects of a therapeutic intervention for foster preschoolers on diurnal cortisol activity. *Psychoneuroendocrinology*, *32*, 892–905.

Fleiss, J., Levin, B., & Paik, M. (2003). (3rd ed.). Hoboken, NJ: Wiley.

Francis, D. D., Caldji, C., Champagne, F., Plotsky, P. M., & Meaney, M. J. (1999). The role of corticotropin-releasing factor-norepinephrine systems in mediating the effects of early experience on the development of behavioral and endocrine responses to stress. *Biological Psychiatry*, *46*, 1153–1166.

Gunnar, M. R., Brodersen, L., Nachmias, M., Buss, K., & Rigatuso, J. (1996). Stress reactivity and attachment security. *Developmental Psychobiology*, *29*, 191–204.

Gunnar, M. R., Fisher, P. A., & the Early Experience, Stress and Prevention Network. (2006). Bringing basic research on early experience and stress neurobiology to bear on preventive interventions for neglected and maltreated children. *Development and Psychopathology*, *18*, 651–677.

Gunnar, M. R., & Quevedo, K. (2007). The neurobiology of stress and development. *Annual Review of Psychology*, *58*, 145–173.

Gunnar, M. R., & Vazquez, D. M. (2001). Low cortisol and a flattening of expected daytime rhythm: Potential indices of risk in human development. *Development and Psychopathology*, *13*, 515–538.

Hane, A. A., & Fox, N. A. (2006). Ordinary variations in maternal caregiving influence human infants' stress reactivity. *Psychological Science*, *17*, 550–556.

Hart, D., Burock, D., London, B., Atkins, R., & Bonilla-Santiago, G. (2005). The relation of personality types to physiological, behavioural, and cognitive processes. *European Journal of Personality*, *19*, 391–407.

Hart, J., Gunnar, M., & Cicchetti, D. (1995). Salivary cortisol in maltreated children: Evidence of relations between neuroendocrine activity and social competence. *Development & Psychopathology*, *7*, 11–26.

Heim, C., Ehlert, U., & Hellhammer, D. H. (2000). The potential role of hypocortisolism in the pathophysiology of stress-related bodily disorders. *Psychoneuroendocrinology*, *25*, 1–35.

Hertsgaard, L., Gunnar, M., Erickson, M. G., & Nachmias, M. (1995). Adrenocortical responses to the strange situation in infants with disorganized/disoriented attachment relationships. *Child Development*, *66*, 1100–1106.

Kraemer, H. C., Wilson, G. T., Fairburn, C. G., & Agras, W. S. (2002). Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry*, *59*, 877–884.

Levine, S. (1957). Infantile experience and resistance to physiological stress. *Science*, *126*, 405–406.

Liang, K.-Y., & Zeger, S. (1986). Longitudinal data analysis for discrete and continuous outcomes using Generalized Linear Models. *Biometrika*, *84*, 3–32.

Little, R., & Rubin, D. (2002). *Statistical analysis with missing data*. Hoboken, NJ: Wiley.

Liu, D., Diorio, J., Tannenbaum, B., Caldji, C., Francis, D., Freedman, A., et al. (1997). Maternal care, hippocampal glucocorticoid receptors, and hypothalamic-pituitary-adrenal responses to stress. *Science*, *277*, 1659–1662.

Loney, B. R., Butler, M. A., Lima, E. N., Counts, C. A., & Eckel, L. A. (2006). The relation between salivary cortisol, callous-unemotional traits, and conduct problems in an adolescent, non-referred sample. *Journal of Child Psychology and Psychiatry*, *47*, 30–36.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Lupien, S. J., King, S., Meaney, M. J., & McEwen, B. S. (2000). Child's stress hormone levels correlate with mother's socioeconomic status and depressive state. *Biological Psychiatry*, *48*, 976–980.

Luthar, S. S., Cicchetti, D., & Becker, B. (2000). The construct of resilience: A critical evaluation and guidelines for future work. *Child Development*, *71*, 543–562.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, *7*, 83–104.

McBurnett, K., King, J., & Scarpa, A. (2003). The HPA and the development of aggressive, antisocial, and substance abuse disorders. In D. Cicchetti & E. Walker

(Eds.), *Neurodevelopmental mechanisms in psychopathology* (pp. 324–344). New York: Cambridge University Press.

McBurnett, K., Lahey, B. B., Rathouz, P. J., & Loeber, R. (2000). Low salivary cortisol and persistent aggression in boys referred for disruptive behavior. *Archives of General Psychiatry, 57*, 38–43.

Meaney, M. J. (2001). Maternal care, gene expression, and the transmission of individual differences in stress reactivity across generations. *Annual Review of Neuroscience, 24*, 1161–1192.

Nachmias, M., Gunnar, M., Mangelsdorf, S., Parritz, R. H., & Buss, K. (1996). Behavioral inhibition and stress reactivity: The moderating role of attachment security. *Child Development, 67*, 508–522.

Pajer, K., Gardner, W., Rubin, R. T., Perel, J., & Neal, S. (2001). Decreased cortisol levels in adolescent girls with conduct disorder. *Archives of General Psychiatry, 58*, 297–302.

Popma, A., et al. (2006). Hypothalamus pituitary adrenal axis and autonomic activity during stress in delinquent male adolescents and controls. *Psychoneuroendocrinology, 31*, 948–957.

Raine, A. (1993). Features of borderline personality and violence. *Journal of Clinical Psychology, 49*, 277–281.

Raine, A., Venables, P. H., Dalais, C., Mellingen, K., Reynolds, C., & Mednick, S. A. (2001). Early educational and health enrichment at age 3–5 years is associated with increased autonomic and central nervous system arousal and orienting at age 11 years: Evidence from the Mauritius Child Health Project. *Psychophysiology, 38*, 254–266.

Robinson, E. A., & Eyberg, S. M. (1981). The dyadic parent–child interaction coding system: Standardization and validation. *Journal of Consulting and Clinical Psychology, 49*, 245–250.

Rutter, M. (2008). Introduction. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*, 2485–2489.

Rutter, M., Moffitt, T. E., & Caspi, A. (2006). Gene-environment interplay and psychopathology: Multiple varieties but real effects. *Journal of Child Psychology and Psychiatry, 47*, 226–261.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147–177.

Schreiber, J. E., Shirtcliff, E., Van Hulle, C., Lemery-Chalfant, K., Klein, M. H., Kalin, N. H. et al. (2006). Environmental influences on family similarity in afternoon *cortisol* levels: Twin and parent-offspring designs. *Psychoneuroendocrinology, 31*, 1131–1137.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420–428.

Snoek, H., Van Goozen, S., Matthys, W., Buitelaar, J. K., & Van Engeland, H. (2004). Stress responsivity in children with externalizing behavior disorders. *Development and Psychopathology, 16*, 389–406.

Sobel, M. E. (1990). Effect analysis and causation in linear structural equation models. *Psychometrika, 55*, 495–515.

Thompson, L. A., & Trevathan, W. R. (2008). Cortisol reactivity, maternal sensitivity, and learning in 3-month-old infants. *Infant Behavior and Development, 31*, 92–106.

Tremblay, R. E. (2008). Understanding development and prevention of chronic physical aggression: Towards experimental epigenetic studies. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*, 2613–2622.

van de Wiel, N. M. H., van Goozen, S. H. M., Matthys, W., Snoek, H., & van Engeland, H. (2004). Cortisol and treatment effect in children with disruptive behavior disorders: A preliminary study. *Journal of the American Academy of Child and Adolescent Psychiatry, 43*, 1011–1018.

van Goozen, S., Fairchild, G., Snoek, H., & Harold, G. T. (2007). The evidence for a neurobiological model of childhood antisocial behavior. *Psychological Bulletin, 133*, 149–182.

van Goozen, S. H., Matthys, W., Cohen-Kettenis, P. T., Buitelaar, J. K., & van Engeland, H. (2000). Hypothalamic-pituitary-adrenal axis and autonomic nervous system activity in disruptive children and matched controls. *Journal of the American Academy of Child and Adolescent Psychiatry, 39*, 1438–1445.

van Goozen, S. H., Matthys, W., Cohen-Kettenis, P. T., Gispen-de Wied, C., Wiegant, V. M., & van Engeland, H. (1998). Salivary cortisol and cardiovascular activity during stress in oppositional-defiant disorder boys and normal controls. *Biological Psychiatry, 43*, 531–539.

Webster-Stratton, C. (1985). Comparisons of behavior transactions between conduct-disordered children and their mothers in the clinic and at home. *Journal of Abnormal Child Psychology, 13*, 169–183.

Webster-Stratton, C. (1990). *Incredible years dinosaur social skills and problem-solving curriculum*. Seattle, WA: Incredible Years Press.

Webster-Stratton, C., & Hancock, L. (1998). Training for parents of young children with conduct problems: Content, methods, and therapeutic processes. In C. E. Schaefer & J. M. Beriesmeister (Eds.), *Handbook of parenting training* (pp. 98–152). New York: Wiley.

Webster-Stratton, C., Reid, M., & Hammond, M. (2004). Treating children with early-onset conduct problems: Intervention outcomes for parent, child, and teacher training. *Journal of Clinical Child & Adolescent Psychology, 33*, 105–124.

Williams, J. B., Gibbon, M., First, M. B., Spitzer, R. L., Davies, M., Borus, J., et al. (1992). The Structured Clinical Interview for DSM–III–R (SCID): II. Multisite test–test reliability. *Archives of General Psychiatry, 49*, 630–636.

Young, E. A., Vazquez, D., Jiang, H., & Pfeffer, C. R. (2006). Saliva cortisol and response to dexamethasone in children of depressed parents. *Biological Psychiatry, 60*, 831–836.

Zeger, S., Liang, K.-Y., & Albert, P. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics, 44*, 1049–1060.

Zuckerman, M. (1979). *Sensation seeking: Beyond the optimal level of arousal*. Hillsdale, NJ: Erlbaum.

Child Development, January/February 2010, Volume 81, Number 1, Pages 306–325

Early-Childhood Poverty and Adult Attainment, Behavior, and Health

Greg J. Duncan
University of California Irvine

Kathleen M. Ziol-Guest
Institute for Children and Poverty

Ariel Kalil
University of Chicago

This article assesses the consequences of poverty between a child's prenatal year and 5th birthday for several adult achievement, health, and behavior outcomes, measured as late as age 37. Using data from the Panel Study of Income Dynamics (1,589) and controlling for economic conditions in middle childhood and adolescence, as well as demographic conditions at the time of the birth, findings indicate statistically significant and, in some cases, quantitatively large detrimental effects of early poverty on a number of attainment-related outcomes (adult earnings and work hours). Early-childhood poverty was not associated with such behavioral measures as out-of-wedlock childbearing and arrests. Most of the adult earnings effects appear to operate through early poverty's association with adult work hours.

Some 4.2 million infants, toddlers, and preschoolers lived in poverty in the United States in 2007. For a single mother with two children, this meant that total income was less than $16,705; many poor families had income well below that amount (U.S. Census Bureau, 2008). Poverty and its attendant stressors have the potential to shape the neurobiology of the developing child in powerful ways, which may lead directly to poorer outcomes later in life. Poverty in early childhood can also affect adult attainment, behavior, and health indirectly through parents' material and emotional investments in children's learning and development.

The sensitivity of early childhood to environmental influences has been demonstrated in a wide range of infant, toddler, and preschooler intervention studies. Many descriptive studies show that, relative to nonpoor children, poor children will be less successful in school and, as adults, in the labor market, have poorer health, and be more likely to commit crimes and engage in other forms of problem behavior (see Holzer, Schanzenbach, Duncan, & Ludwig, 2007, for a review of these studies). Despite these associations, it is far from clear to what extent poverty *itself* is the cause of these differences. Our primary goal in this study is to obtain relatively unbiased estimates of the total effects association between early-childhood poverty and adult attainment, behavior, and health. Extending the work of Duncan, Brooks-Gunn, Yeung, and Smith (1998), we use the most recent data from the Panel Study of Income Dynamics (PSID) to examine the long-run (i.e., as late as age 37) impacts of low income early in life, net income later in childhood, and other correlated family factors surrounding a child's birth.

Background

Emerging evidence from human and animal studies highlights the critical importance of early childhood for brain development and for setting in place the structures that will shape future cognitive, social, emotional, and health outcomes (Sapolsky, 2004; Shonkoff & Phillips, 2000). How poverty early in childhood might affect these structures has been

---

We thank William Dickens, Rob Dugger, Mimi Engel, Michael Foster, Rucker Johnson, Ann Siegal, Aaron Sojourner, Sergio Urzua, Sara Watson, and participants of the University of Michigan's National Poverty Center conference "Long-Run Impact of Early Life Events" and of seminars at the Brookings Institution, Brown University, Cornell University, and the University of Oxford for comments on related drafts. This project was funded in part by the Partnership for America's Economic Success. Ziol-Guest acknowledges the Robert Wood Johnson Foundation Health and Society Scholars program at Harvard University for its financial support.

Correspondence concerning this article should be addressed to Greg J. Duncan, Department of Education, University of California Irvine, Irvine, CA 92697. Electronic mail may be sent to gduncan@uci.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8101-0020

a matter of considerable research. Economic models of skill acquisition illustrate the potential importance of early environmental conditions on adult cognitive and noncognitive skills (Knudsen, Heckman, Cameron, & Shonkoff, 2006). Among other things, low income during early childhood may limit parents' ability to purchase adequate-quality health care or education during children's formative years.

Complementary studies in psychology and social epidemiology illustrate that both in utero environments and early-childhood experiences have long-run impacts on adult physical and mental health (Barker, Eriksson, Forsen, & Osmond, 2002; Danese, Pariante, Caspi, Taylor, & Poulton, 2007; Poulton & Caspi, 2005). Epidemiologists have suggested that early-childhood stressors related to low income could alter or dysregulate biological systems, with adverse implications for future health (Godfrey & Barker, 2000). Psychologists posit that poverty and economic insecurity undermine parents' mental health and parenting behavior. In animal models, optimal "mothering" behavior in critical periods of early development is associated with lifelong stress reactivity and cognitive strength (Sap, 2004).

Duncan and Brooks-Gunn (1997) were the first to take a broad look at the possible longer run consequences of early-childhood poverty. Twelve groups of researchers working with 10 different nonexperimental but longitudinal data sets estimate longitudinal models of early-childhood income effects on later attainment, behavior, and health. On the whole, the results suggest that family income has substantial, albeit selective associations with children's subsequent attainments.

First, family income had consistently larger associations with measures of children's cognitive ability and achievement than with measures of behavior, mental health, and physical health. Second, family economic conditions in early childhood appeared to be more important for shaping ability and achievement than did family economic conditions during adolescence. And third, the association between income and achievement appeared to be nonlinear, with the biggest impacts at the lowest levels of income.

Why the Timing of Income May Matter

The present study builds on the key conclusion from Duncan and Brooks-Gunn (1997) that family economic conditions in early childhood appear to matter more for shaping later development than economic conditions during adolescence. Developmental theory suggests that given the nature of developmental tasks, sensitivity to change, and interactions with the environment, early childhood is a developmental period that may be especially sensitive to environmental conditions affected by family income (Shonkoff & Phillips, 2000). Moreover, early courses of development may reach well into adulthood. Waddington (1957) has described development as proceeding along the branches of a tree—although changes in developmental trajectories can occur at any point at which a new branch is formed, the ability of the individual to alter his or her developmental course substantially becomes increasingly difficult over time. These themes are reflected in economic, psychological, and neurobiological perspectives on the importance of early childhood.

Cunha, Heckman, Lochner, and Masterov (2005) proposed an economic model of development in which preschool cognitive and socioemotional capacities are key ingredients for human capital acquisition during the school years. In their model, "skill begets skill" as early capacities can affect the productivity of school-age human capital investments. Economic deprivation in early childhood could create disparities in school readiness and early academic success that persist or widen over the course of childhood. The theory proposed by Heckman and colleagues is novel in economics for its developmental perspective: Typically, the economics literature has ignored the notion that the effects on children's development of economic conditions may depend upon childhood stage and instead focuses on the role of "permanent" income, with the assumption that families anticipate bumps in their life-cycle paths and can save and borrow freely to smooth their consumption across these bumps (Blau, 1999).

Developmental psychology clearly stresses the importance of understanding children's distinct developmental stages (Bronfenbrenner & Morris, 1998). In the context of poverty studies, the greater malleability of children's development and the overwhelming importance of the family (as opposed to school or peer contexts) for preschoolers lead us to expect that family income in early childhood may be much more important for shaping children's ability and achievement than conditions later in childhood (Bronfenbrenner & Morris, 1998; Shonkoff & Phillips, 2000). For example, to the extent that poverty increases mothers' psychological stress or harsh parenting behaviors (Yeung, Linver, & Brooks-Gunn, 2002), this will be especially important during early childhood given the primacy of sensitive mother–child interactions for the development of young children's emotion regulation (Waters & Sroufe, 1983). Mastery of emotion regulation in early childhood can have long-run impacts on children's achievement, behavior, and health (Fox, 1994). Similarly, to the extent that cognitively enriching early home environments lay the groundwork for success in preschool and beyond, parents' ability to purchase books, toys, and enriching activities during this stage of development is paramount.

As its name suggests, the "fetal origins" hypothesis also points to a sensitive period in early childhood. In this case, the in utero environment is critical. This hypothesis posits an in utero programming process whereby stimulants and insults during this critical period of development have long-lasting implications for physiology and disease risk (Barker et al., 2002; Godfrey & Barker, 2000; Hertzman, 1999). Low income during the prenatal period may be associated with fetal undernutrition, low birth weight, or slow growth in the first 2 years of life. A pattern of small size at birth and low body mass index (BMI) at age 2, followed by rapid weight gain after age 2, is a risk factor for the development of insulin resistance and a disproportionately high fat mass in relation to muscle mass (Barker, Osmond, Forsén, Kajantie, & Eriksson, 2005; Barker et al., 2002). Low income is also associated with food insecurity, which has been associated with overweight and obesity in childhood, adolescence, and adulthood, especially among females (Frongillo, 2003). Overweight and obesity, in turn, can be physically debilitating, which could lead to a negative cycle of depression, overeating, or stress (Stunkard, Faith, & Allison, 2003).

Methods for Assessing Causal Impacts of Poverty

Researchers generally do not dispute simple correlations between income and child developmental outcomes. However, there is much controversy about whether these income effects are causal. The key estimation problems in assessing the causal impact of family income on child well-being are twofold: timing of measurement and omitted-variable bias. Theory suggests that the development of children's cognitive and social skills is a time-consuming process. Attainments in, say, adolescence, may well be a product of economic conditions not only in adolescence but also in early and middle childhood and possibly during the prenatal period as well (Barker, 1998). Estimates of models of income effects that measure income concurrently with the child outcomes risk bias if income is volatile across childhood. As there is abundant evidence that income is indeed volatile (Duncan, 1988), a longitudinal perspective on the role of income in shaping child well-being appears crucial.

Even supposing that income is measured well across the entire period of childhood, it is difficult to isolate the causal impact of income, as many factors might jointly determine family income and child development and parental cognitive ability is a prime example (Rowe & Rodgers, 1997). Parents with higher cognitive ability are usually more successful in the labor market. At the same time, they are more likely to provide a higher quality learning environment for their children, regardless of how much money they may be spending on books or computers. Some studies find large reductions in the estimated impacts of income once adjustments for omitted-variable bias are implemented (Blau, 1999; Mayer, 1997).

State or national policies sometimes change in ways that provide researchers with opportunities to relate exogenous, policy-induced changes in family income to child well-being. Such is the case with the study by Dahl and Lochner (2005), who took advantage of the fact that the United States increased the generosity of its Earned Income Tax Credit (EITC) program during the 1990s. The EITC provides a refundable tax credit to low-income working families. The maximum size of the annual credit is now quite substantial—$4,824—and it increased by about $2,300 in the middle 1990s. Dahl and Lochner estimated that a $3,000 increase in family income in early and middle childhood boosts reading achievement by about one tenth of a standard deviation and math achievement by about half that amount. These effects were two to three times as large, however, for children of non-White, unmarried, and less educated mothers, which corresponds to another key conclusion in Duncan and Brooks-Gunn (1997)—namely, that income effects are nonlinear such that increases in income matter more for lower income children than their higher income counterparts.

The different disciplines' approaches to these questions each have their own merits and limitations. Developmental research articles on the impacts of poverty, although often paying close attention to the processes or pathways linking poverty with developmental outcomes, typically employ only a handful of family-based demographic covariates, opening the door for a host of omitted-variable biases. Research in economics, which generally takes a "total effects" (reduced-form) approach, typically opts to exploit some kind of natural experiment—siblings differences within families or temporal or areal variation in income induced by policy (e.g., EITC expansion) or macroeconomic (e.g., unemployment differences) fluctuations. These approaches have stronger claims to internal validity (reduced bias) but, concentrating as they often do on narrow subpopulations, often sacrifice external validity. Such studies also often yield very imprecise estimates. And no prior study has been able to link income-based economic conditions in early childhood to outcomes measured well into adulthood.

We use data from a long-running national survey to surmount some of these limitations. As in many studies in economics, we aim to estimate the total effects (as opposed to a mediational model) of early-childhood poverty and adult attainment, behavior, and health. By taking a population-based approach, we minimize threats to external validity. And, in contrast to many developmental studies, we employ an unusually rich and powerful array of covariates to control for omitted-variable biases. In particular, our estimates of the impact of early-childhood poverty on adult attainment, behavior, and health include controls for income in middle childhood and adolescence. It is very difficult to think up omitted-variable bias stories involving early income that would not be controlled in large measure with the inclusion of income later in childhood.

This study draws on national data from the PSID, the longest-running longitudinal study of household income in the United States, to estimate linkages between income early in childhood and adult outcomes. Ours is the first study to link high-quality income data across the entire childhood period with adult outcomes measured as late as age 37. Our strategy is to measure income in every year of a child's life from the prenatal period through age 15, distinguishing income early in life (prenatal through 5th year) from income in middle childhood and adolescence. Our analyses relate an array of adult achievement, social assistance, health and behavior measures to these childhood stage-specific measures of income, plus a host of relevant demographic control variables measured around point of the child's birth. The wide range of adult outcomes we consider include educational attainment, earnings, work hours, receipt of food stamps and cash assistance, nonmarital childbearing, crime, and mental and physical health.

Method

Sample

We use 1968–2005 data from the PSID, which has followed a nationally representative sample of families and their children since 1968 (http://psidonline.isr.umich.edu). Our general strategy is to select children observed in the PSID between their prenatal year and at least age 25, although a number of our outcomes are drawn from the 2005 interview, in which individuals ranged from ages 30 to 37.

Our target study sample consists of the 2,599 individuals born to a sample member into PSID households between 1968 and 1975, who were thus between ages 30 and 37 in 2005. Sample losses to nonresponse are substantial, prompting us to use the PSID's attrition-adjusted weights in all of our analyses. We required that these individuals have complete data on control variables measured around the time of their birth (defined next), leaving 93% of the birth sample ($n = 2{,}414$). We further required that these individuals be in response families in at least 12 of the 17 years from prenatal year to age 15, a restriction that eliminated 479, or 18% of the original sample. Of the remaining 2,120, 1,589 participated in the PSID in adulthood and had nonmissing data on at least one outcome and a PSID attrition-adjusted weight. Sample sizes vary by outcome owing to when it was measured, and in some cases (distress and body mass) only information was collected from individuals who were "heads" or "wives" in PSID households.

Childhood Income

We used the PSID's high-quality edited measure of annual total family income, inflated to 2005 levels using the Consumer Price Index (CPI). We averaged these annual income measures across three periods: the prenatal year through the calendar year in which the child turned 5, ages 6–10, and ages 11–15. With income reported for calendar years and conceptions occurring continuously, there was some imprecision in matching income to the prenatal year and beyond. If a child was born prior to July 1, we took the prenatal year to be the prior calendar year. If the birth was after July 1, then the prenatal year was considered to be the year in which the birth occurred. Similarly, we defined "under age 6" as the last calendar year before the child's sixth birthday. Thus defined, our "early-childhood" period consists of seven calendar years.

To account for a possible differential impact of increments to low as opposed to higher family income, we allowed the coefficients on average income within each childhood period to have distinct linear effects for average incomes up to $25,000 and for incomes $25,000 and higher. Considerable experimentation confirmed the utility of the $25,000 "knot" in the case of the adult outcomes most sensitive to variations in early income; lower income knots reduced sample sizes for, and therefore the precision of, the low-income effect estimates; higher thresholds typically produced noteworthy reductions in the low-income segment coefficients.

Adult Outcomes

Dependent variables in our analyses spanned achievement, health, and behavioral domains. Years of completed schooling are based on the most recent report of schooling available in the data. In all cases, the report was taken when the individual was at least 22 years old and in most cases the individual was at least 25.

Adult work hours and natural logarithm of the child's adult earnings were gleaned from all available annual reports of earned income and work hours reported by or for the child when the child was aged 25 or older. As with childhood income, we inflated the dollar values of earnings to 2005 price levels using the CPI. To adjust earnings for age and calendar year, we regressed all of our yearly earnings observations on sets of dummy variables measuring the age of the respondent in the given year, and the calendar year of measurement. We then generated residuals from this regression for each sample individual's earnings observations and averaged these residuals across all of the yearly earnings observations that a given individual generated. We centered these average residuals around the sample mean by adding them to the overall sample mean earnings. We repeated this process to adjust work hours for age and year effects. As a final step in the case of earnings, we took the natural logarithm.

Food stamp and Aid to Families with Dependent Children/Temporary Assistance to Needy Families (AFDC/TANF) receipt are measured at the household level and are taken from all available surveys when the child was aged 25 or older. Although some adults under age 25 are program recipients, it was difficult to assign transfer income sources to individual household members prior to around age 25. We created calendar-year values of both programs and inflated the values to 2005 price levels using the CPI. Like average annual earnings, we adjust for age and calendar-year effects by regressing all food stamp and AFDC/TANF values on age and calendar-year dummies, obtained the residuals, and calculated the average residuals and mean values across all available years for a given individual. We estimate food stamp models for the entire sample but AFDC/TANF models only for the females.

Our measure of poor overall health was based on the most recent response to the question "I have a few questions about your health, including any serious limitations you might have. Would you say your health in general is excellent, very good, good, fair, or poor?" Individuals are considered in poor health if they responded that their health was either fair or poor. Although the results are not reported in the tables, we also ran our models using a continuous measures of health-assigning integer values of 1–5 for these respective categories and with an alternative scaling of *excellent* = 100, *very good* = 85, *good* = 70, *fair* = 30, and *poor* = 0. Our regression results are robust to this specification. Because we used the most recently available report of self-rated health, our regression analyses of this outcome also include calendar-year dummy variables for when the individual's report was taken.

Our measure of adult BMI was calculated based on reports in the 2005 survey of heads and wives of their weight in pounds and their height in feet and inches. We calculate BMI using the following formula:

$$\frac{\text{Weight} \times 703}{\text{Height}^2},$$

where weight is measured in pounds and height is measured in inches. We follow convention and define "overweight" as a BMI ≥ 25 (Centers for Disease Control and Prevention, 2003).

Our measure of psychological distress was based on responses to a 2003 administration to heads and wives of the K-6 NonSpecific Psychological Distress Scale, developed by Ronald Kessler. It includes six items, ranging from *all of the time* = 4, *most of the time* = 3, *some of the time* = 2, *a little of the time* = 1, and *none of the time* = 0. The questions are: "Now, I am going to ask you some questions about feelings you may have had over the past 30 days." In the past 30 days, about how often did you feel: (a) so sad nothing could cheer you up? (b) nervous? (c) restless or fidgety? (d) hopeless? (e) that everything was an effort? and (f) worthless? The scores are summed; a score of 13 or higher is considered to be the threshold for the clinically significant range of the distribution of nonspecific psychological distress, which we refer to here as ''high distress.'' As with the general health measure, we also analyzed a continuous measure of distress and found the same general patterns.

The crime outcomes consisted of responses to questions asked in the 1995 interviewing wave regarding past arrests and time in jail. A dichotomous ''arrest'' outcome was coded for an affirmative response to the question: ''Not counting minor traffic offenses (has he/has she/have you) ever been booked or charged for breaking a law?'' A dichotomous ''jail'' outcome was coded for an affirmative response to the question: ''(Has he/Has she/Have you) ever spent time in a corrections institution like a jail, a prison or a youth training or reform school?'' Owing to the infrequency of arrest and incarceration among females, this analysis was restricted to males.

Finally, our nonmarital birth outcome is based on a dichotomous indicator of whether the individual (females only) reported a nonmarital birth prior to her 21st birthday in the PSID's fertility and marital histories.

Control Variables and Regression Procedures

To avoid attributing to income what should be attributed to correlated determinants of both childhood income and our outcomes of interest, we included the following control variables in all of our regressions: (a) dummy variables for seven of the eight birth years, (b) race (Black, other, with White the reference category), (c) child's sex (male = 1), (d) whether the child's parents were married and living together at the time of the birth, (e) the age of the mother at the time of the birth, (f) whether the child lived in the South at the time of the birth, (g) the total number of siblings born to the child's mother, (h) whether the child was the firstborn to his or her mother, (i) years of completed schooling of the household head (usually the father in two-parent households, the mother in single-parent households) of the child's household in the birth year, and (j) the head's score on a sentence completion test administered in the 1972 interviewing wave.

Additionally, we include several variables taken in the earliest years of the PSID and use the survey response from the year closest to the child's birth. All but the first of these measures are collected from the household head. First, we include the response to an interviewer observation regarding the cleanliness of the respondent's dwelling. Responses ranged from 1 (*very clean*) to 5 (*dirty*); therefore, higher values represent perceptions of less clean home environments. Dunifon, Duncan, and Brooks-Gunn (2001) found significant links between this measure and children's completed schooling. Second, we include an index of parental expectations for children that reflects expectations for their own as well as their children's futures. Components of this scale include responses to statements such as having explicit plans for children's education and jobs. Third, we include two motivational measures: (a) an assessment of the importance of challenge versus affiliation and (b) a measure of an individual's sense of personal control. Both of these measures are associated with long-run labor market success (Duncan & Dunifon, 1998). The challenge versus affiliation measure was created by averaging responses to two questions ''Would you liked to have more friends, or would you like to do better at what you try?'' and ''Would you prefer a job where you had to think for yourself, or one where you work with a nice group of people?'' A score of 1 indicates preference for challenge, whereas those who preferred to have more friends or work with nice people received scores of 0. Sense of personal control was obtained through three questions: ''Have you usually felt pretty sure that your life will work out the way you want it to, or have there been more times when you haven't been sure about it?'' ''When you make plans ahead, do you usually get to carry out things the way you expected, or do things usually come up to make you change your plans?'' and ''Would you say that you nearly always finish things once you start them, or do you sometimes have to give up before they are finished?'' Responses were scored 1 for a positive response, 0.5 for an equivocal response, and 0 for a negative response. Finally, we included a measure of risk avoidance, which ranges from 0 to 9 capturing avoiding risk on everyday things including having at least one car in good condition, all cars are insured, using seatbelts, health insurance or free medical care, family does not smoke, and family has some savings.

All of our regressions were run in STATA SE 9.0 standard error, use the PSID's weights to adjust for differential sampling fractions and attrition, and adjust for origin–family clustering on the mother using Huber–White methods. Experimentation with both weighted and unweighted regression estimates revealed marked differences in coefficients on early-childhood income in several cases, which led to using the weights in the regression analyses—a point discussed further in the Results section.

Continuous outcomes (ln earnings, work hours, and completed schooling) were analyzed with ordinary least squares (OLS), measures with substantial concentrations of zeroes (food stamp and AFDC/TANF receipt) were analyzed with Tobit regressions, and dichotomous outcomes (poor health, high distress, ever arrested, ever jailed, and nonmarital birth before age 21) were analyzed with logistic regression. The arrest and incarceration models are only run on males, whereas the AFDC/TANF and nonmarital childbearing models are only run on females. To facilitate their interpretation, logistic regression coefficients and standard errors are expressed in the tables in the form of marginal effects (and their associated standard errors), computed using the MFX command in STATA. These show the change in the probability (as opposed to the odds) of the outcome occurring associated with a unit change in the given independent variable.

Results

Sample Description

Case counts, means, and standard deviations of each of our outcome variables are provided in Appendix A. These summary statistics are weighted using the weight for the year in which the outcome was measured (high distress, BMI, overweight, arrest, and incarceration), or the most recent weight of the PSID (completed schooling, earnings, work hours, food stamp, and AFCD/TANF receipt and nonmarital childbearing). Descriptive statistics are presented for both the overall sample and for children whose prenatal-to-age-5 incomes averaged: (a) below the official poverty line, (b) between 1 and 2 times the poverty line, and (c) more than twice the poverty line. The final column of Appendices A and B provide information on the statistical significance (at $p < .05$ or below, two-tailed test) of the mean differences across the three groups.

Appendix A shows striking differences in adult outcomes depending on whether childhood income prior to age 6 was below, close to, or well above the poverty line. Compared with children whose families had incomes of at least twice the poverty line during their early childhood, poor children complete 2 fewer years of schooling, work 451 fewer hours per year, earn less than half as much, received $826 per year more in food stamps as adults, and are more than twice as likely to report poor overall health or high levels of psychological distress. Further, poor children have BMIs that are 4 points higher than those well above the poverty line, and are almost 50% more likely to be overweight as adults. Poor males are twice as likely to be arrested and for females, poverty is associated with a $200 annual increase in cash assistance, and a sixfold increase in the likelihood of bearing a child out of wedlock prior to age 21.

Appendix B reports the weighted descriptive statistics of the childhood period income measures and control variables for the total sample, as well as by poverty status in early childhood. Not surprisingly, children with average annual incomes below poverty in the earliest period have lower average income in all three periods compared with the other two groups. Additionally, the poorest children are less likely to be White and born into an intact family, and more likely to be born in the South, have younger mothers, more siblings, household heads with lower test scores and educational attainment, homes rated as dirtier by interviewers, lower parental expectations, and household heads who report less preference for challenge versus affiliation, less personal control, and less risk avoidance compared with their higher income counterparts.

Multivariate Results

To motivate the estimates from our full regression models, we present in Table 1 standardized coefficient results from a series of descriptive regressions. Across the columns of the first row (Model 1), each outcome is regressed only on a measure of 17-year average childhood income. For ease of calculation of standardized coefficients for our health outcomes, we estimate OLS regressions using continuous measures of general health and distress. Further, to calculate the standardized logistic regression coefficients, we use the following equation from Menard (2004): $b^* = (b)(s_x)(R)/s_{logit(\hat{y})}$, where b is the logit coefficient, s_x is the standard deviation of x, R is the square root of the R^2 from the logit regression, and $s_{logit(\hat{y})}$ is the standard deviation of the predicted value of y.

As we express estimates as standardized regression coefficients, entries in the first row amount to simple correlations between income and each of the outcomes. The directions of all of the correlations are as expected—positive for ''good'' outcomes and negative for ''bad'' ones—and significant for all

Table 1
Standardized Regression Coefficients From Various Weighted Models of Childhood Income and Adult Outcomes

	Income measure	Schooling (years)	Annual earnings	Annual work hours	Food stamps	AFDC/ TANF	Health status	Distress	BMI	Arrested	Nonmarital birth
Model 1: No controls; 17-year average childhood income	Prenatal to age 15	.34**	.31**	.15**	−.25**	−.13**	−.13**	−.11**	−.15**	−.17*	−.38**
Model 2: Background controls; 17-year average childhood income	Prenatal to age 15	.14**	.23**	.11**	−.08*	−.08†	−.02	−.06	−.02	−.09	−.26**
Model 3: Background controls; natural logarithm of 17-year average childhood income	Prenatal to age 15	.22**	.37**	.20**	−.24**	−.07	−.08*	−.09	−.09†	−.13†	−.18**
Model 4: Background controls; natural logarithm ofaverage stage-specific childhood income	Prenatal to age 5	.10*	.27**	.20**	−.08†	−.08	.07	.00	−.04	−.11	.02
	Ages 6–10	−.04	.06	−.03	−.12*	.07	−.05	.14†	−.09	.11	.07
	Ages 11–15	.18**	.08	.06	−.06	−.07	−.11†	−.25**	.04	−.17†	−.27**

Note. Ordinary least squares analysis for schooling, earnings, work hours, food stamps, Aid to Families with Dependent Children/Temporary Assistance to Needy Families (AFDC/TANF), health status, distress, and body mass index (BMI); logit analysis for arrests and nonmarital births. Standardized regression coefficients for the logistic regression employ procedures from Menard (2004).
†$p < .10$. *$p < .05$. **$p < .01$.

outcomes. The largest correlations, all in the .3–.4 range, are found for schooling, adult earnings, and nonmarital births.

In the second row (Model 2) results, our extensive set of background control variables, all measured either before or right around the time of birth, is added to each of the regressions. All of the coefficients become smaller (in absolute value); in the case of the three health measures and arrests, the coefficient on childhood income drops below the .05 threshold of statistical significance. Thus, a substantial portion of the simple correlation between childhood income and most adult outcomes can be accounted for by the disadvantageous conditions associated with birth into a low-income household.

To assess whether increments to low income may matter more than increments to the incomes of children growing up in middle-class or affluent families, we regressed the adult outcomes on the natural logarithm of the 17-year average childhood income plus background controls. Whereas our first two models assumed that, say, a $5,000 increment to a poor family's income had the same beneficial effect on a child's adult outcomes as a $5,000 increment to an affluent family's income, the logarithmic transformation assumes equal percentage effects. So, for example, the logarithmic model presumes that a 50% (and $10,000) increase in average childhood income from $20,000 to $30,000 has the same effect as the 50% (but $50,000) increase from $100,000 to $150,000. Higher standardized coefficients (in absolute value) in logarithmic as opposed to linear models would suggest that dollars matter more for the developmental outcomes of children reared in lower than higher income households.

As shown in the third (Model 3) row of Table 1, the standardized coefficients for the logarithmic models are uniformly higher than coefficients for the linear models (Model 2) in the case of achievement outcomes: completed schooling, earnings, work hours, and food stamps. Health outcomes and AFDC/TANF receipt are little affected by the change, whereas coefficients increase only slightly for arrests and actually fall for nonmarital births. The pattern of results for these adult outcomes is broadly consistent with the conclusions that Duncan and Brooks-Gunn (1997) reached for childhood outcomes: Income appears to be more predictive of achievement-related than either behavior or health outcomes.

To address the issue of the childhood stage-specificity of income effects, the final (Model 4) regressions in Table 1 replace the single 17-year average log childhood income measure with three stage-specific measures of log income. All background controls are included in these models. With each childhood stage accounting for approximately one third of childhood, we would expect that the three coefficients should (approximately) sum to the all-childhood coefficient presented in Model 3. If childhood income mattered equally across all three stages, the three coefficients should be roughly the same size and about one third the magnitude of the Model 3 coefficients.

In the case of adult earnings and work hours, early-childhood income appears to matter much more than later income. For work hours, the standardized coefficient on prenatal-to-age 5 log average income (.20) is just as large as the Model 3 coefficient on all-childhood log average income, suggesting little role for income beyond age 5. Early income also has a statistically significant coefficient ($p < .05$) in the case of completed schooling, but in this case adolescent income has a considerably larger standardized coefficient. Adult health outcomes and nonmarital births have strong associations with childhood income, but only with childhood income during adolescence.

Taken together, the descriptive regression results shown in Table 1 suggest that both childhood stage and outcome domain matter for understanding links between childhood income and adult success. Moreover, many of the adult outcomes appear to be more sensitive to increments to low as opposed to middle-class or high family incomes.

These lessons are incorporated into our main regression models, shown in Table 2 for achievement outcomes and program participation, Table 3 for health outcomes, and Table 4 for behavioral outcomes. Coefficients and standard errors for the childhood-stage-specific income variables are presented first. For each stage's income, two coefficients are presented, the first reflecting the estimated effect of an additional $10,000 of annual income in the given stage for children whose income in that stage averaged < $25,000 and the second reflecting comparable effects for higher income children (all three sets of income variables, plus other controls, are included in all regressions). The column labeled "different slopes" reports results from a statistical test of the null hypothesis of equal within-period slopes. The final row shows results for a test of equality of all three < $25,000 segment slopes. In contrast to Table 1, coefficients in Tables 2–4 are unstandardized and therefore show changes in originally scaled dependent variables associated with $10,000 increments to average childhood stage-specific income.

Table 2 shows that additional income in the prenatal to age 5 period for the lowest-income children is associated with significantly greater adult earnings and work hours, and less food stamp receipt. To ensure that our spline models fit the earnings and work hours data reasonably well, we estimated models with early income represented with a set of dummy variables, and controls for later period incomes (ages 6 through 10 and ages 11 through 15). As shown in Figures 1 and 2, which include 95% confidence bands, the income range in which income responses flattened out was roughly between $20,000 and $30,000, thus supporting our use of a knot at $25,000.

We can illustrate the nature of the income effects of early-childhood income using the natural log earnings regression. The "0.52" coefficient means that, adjusting for income later in childhood and the other control variables listed in Table 2, an additional $10,000 per year of family income between the prenatal year and the child's fifth birthday is associated with an increase in the natural logarithm of adult earnings of 0.52—or 68.2% ($e^{52} = 1.682$). In contrast, increments to early-childhood income for higher-income children (i.e., annual average family incomes above $25,000) are associated with an insignificant 0.05 increment in log earnings. The p value ($p < .05$) reported in the "different slopes" column indicates that the slope for those with incomes less than $25,000 per year in early childhood is significantly different from the slope for those with incomes greater than $25,000 in the same period. Increments to incomes in middle childhood and adolescence are estimated to have nonsignificant impacts on log earnings, even among low-income children. The final row of Table 2 indicates that the three coefficients on the < $25,000 spline slopes across the childhood stages are significantly different from one another at the $p = .08$ level.

Results for work hours are broadly similar to those for earnings—a highly significant estimated impact of early childhood but not later childhood income. In this case, a $10,000 annual increase in the prenatal to age 5 income of low-income families is associated with more than 500 additional work hours per year after age 25. Tobit spline regressions for food stamps for the entire sample suggest that increases in income in all childhood periods for the lowest income children are associated with statistically significant reductions in food stamps. Both the coefficients on food stamps and welfare proved

Table 2

Ordinary Least Squares Spline Regression Models of Childhood Income and Completed Schooling, Adult Earnings and Annual Work Hours, and Tobit Spline Models of Childhood Income and Program Participation

		Years of completed schooling		ln Earnings (ages 25–37)		Annual hours worked (ages 25–37)		Annual food stamp (ages 25–37)		Annual AFDC/TANF Females only (ages 25–37)	
Childhood income (in $10,000)			Different slopes		Different slopes		Different slopes		Different slopes		Different slopes
Average annual	< $25K	0.19 (0.33)	*ns*	0.52* (0.21)	$p < .05$	506.74** (135.39)	$p < .001$	−279.5* (142.08)	$p < .10$	−157.28 (267.04)	*ns*
income (prenatal to age 5)	> $25K	0.03 (0.04)		0.05** (0.02)		20.60* (9.78)		−4.34 (14.17)		12.02 (24.60)	
Average annual	< $25K	0.65** (0.25)	$p < .01$	0.14 (0.14)	*ns*	−60.82 (109.42)	*ns*	−380.30** (128.37)	$p < .01$	419.06† (233.44)	$p < .10$
income (ages 6–10)	> $25K	−0.06 (0.04)		0.01 (0.02)		1.28 (8.36)		3.91 (11.89)		−25.91 (27.44)	
Average annual	< $25K	−0.31 (0.20)	$p < .10$	0.04 (0.12)	*ns*	74.18 (83.33)	*ns*	−221.71* (101.53)	$p < .05$	172.35 (202.78)	*ns*
income (ages 11–15)	> $25K	0.09** (0.03)		0.00 (0.01)		−0.92 (7.56)		−6.07 (9.92)		−37.73 (23.59)	
Other variables											
Black		0.76** (0.24)		0.10 (0.10)		−79.44 (82.40)		368.48** (78.44)		964.63* (155.28)	
Other minority		0.68 (0.57)		−0.12 (0.35)		−255.48 (219.46)		213.80 (134.61)		−312.59 (378.53)	
Child is male		−0.45** (0.13)		0.53** (0.07)		566.54** (48.12)		−191.19** (52.78)		—	—
Child born into intact family		0.63** (0.22)		−0.02 (0.12)		47.66 (80.15)		76.63 (86.08)		−146.97 (172.44)	
Child born in South		−0.16 (0.18)		−0.09 (0.07)		−32.16 (57.65)		−35.53 (58.49)		−511.95** (116.70)	
Age of mother at time of birth		0.04* (0.02)		0.00 (0.01)		1.46 (6.17)		−8.98 (5.85)		−17.56 (12.87)	
Number of siblings		−0.09 (0.06)		0.02 (0.02)		−9.49 (21.89)		20.23 (16.95)		14.67 (35.90)	
Child is firstborn		0.32* (0.15)		0.19* (0.08)		102.51† (58.86)		−79.03 (61.51)		−246.83* (118.84)	
HH head test score (1972)		0.10* (0.04)		−0.01 (0.02)		−8.04 (17.10)		19.06 (15.02)		−33.86 (27.70)	
HH head schooling		0.10** (0.03)		0.02 (0.02)		−6.71 (10.84)		−24.49* (11.75)		17.37 (24.45)	
Observed ''dirty'' home		−0.09 (0.08)		−0.03 (0.04)		12.19 (26.51)		39.42 (29.48)		−12.67 (54.20)	
Parental expectations		0.14* (0.06)		0.00 (0.03)		15.16 (23.21)		−72.21** (24.41)		−187.05** (47.26)	
Challenge versus affiliation		−0.32 (0.23)		−0.04 (0.11)		−6.66 (81.76)		25.80 (80.99)		−9.63 (162.68)	
Personal control		0.26 (0.22)		0.04 (.12)		−78.53 (86.30)		98.89 (89.62)		227.59 (172.27)	
Risk avoidance		0.16** (0.05)		0.04 (0.03)		10.86 (19.70)		−37.43† (19.95)		24.36 (42.44)	
Birth year dummies?		Yes		Yes		Yes		Yes		Yes	
Regression statistics											
R^2		.29		.24		.24		.03		.04	
Number of observations		1,254		1,016		1,042		1,198		601	
p from test of equality of three < $25K spline segments		.04		.08		.01		.73		.38	

Note. Sample consists of Panel Study of Income Dynamics children born between 1968 and 1975. Incomes are in 2005 dollars and childhood incomes are scaled in $10,000. Data in the ''different slopes'' column show p levels of test of equality of within-period < $25K and > $25K slopes. The coefficients and standard errors for the schooling, earnings, and hours are from ordinary least squares (OLS) analysis, and the coefficients and standard errors for food stamps and Aid to Families with Dependent Children/Temporary Assistance to Needy Families (AFDC/TANF) are from Tobit analysis. The AFDC/TANF analysis is only for females. Regressions are weighted and OLS analysis standard errors are corrected for multiple children born in the same household.

†$p < .10$. *$p < .05$. **$p < .01$.

Table 3

Logistic Spline Regression Models of Childhood Income and Poor Health, High Distress, and Overweight; Ordinary Least Squares Spline Regression Model of Childhood Income and Body Mass Index

		Poor health		High distress (2003)		BMI (2005)		Overweight (BMI > 25 in 2005)	
Childhood income (in $10,000)			Different slopes		Different slopes		Different slopes		Different slopes
Average annual	< $25K	−0.009 (0.027)	*ns*	0.011 (0.021)	*ns*	−3.51 (2.45)	*ns*	0.028 (0.113)	*ns*
income (prenatal to age 5)	> $25K	0.003 (0.003)		−0.002 (0.003)		0.04 (0.16)		0.013 (0.014)	
Average annual	< $25K	−0.008 (0.027)	*ns*	0.013 (0.016)	*ns*	0.77 (1.51)	*ns*	−0.130 (0.107)	*ns*
income (ages 6–10)	> $25K	$0.007^{\dagger}$ (0.004)		0.006* (0.002)		−0.12 (0.09)		−0.020* (0.010)	
Average annual	< $25K	0.001 (0.028)	*ns*	−0.026 (0.021)	*ns*	0.34 (1.41)	*ns*	0.128 (0.097)	*ns*
income (ages 11–15)	> $25K	−0.017** (0.003)		−0.006* (0.002)		0.06 (0.09)		0.010 (0.008)	
Other variables									
Black		0.025 (0.025)		−0.024* (0.014)		0.36 (0.83)		−0.039 (0.089)	
Other minority		−0.009 (0.042)		—		−0.77 (0.87)		−0.014 (0.153)	
Child is male		0.014 (0.014)		$-0.024^{\dagger}$ (0.014)		1.73** (0.47)		0.278** (0.043)	
Child born into intact family		0.014 (0.016)		−0.009 (0.021)		−0.09 (0.94)		0.059 (0.093)	
Child born in South		0.007 (0.015)		0.011 (0.013)		−0.20 (0.55)		−0.062 (0.053)	
Age of mother at time of birth		0.001 (0.001)		0.000 (0.001)		0.02 (0.06)		$-0.009^{\dagger}$ (0.006)	
Number of siblings		−0.006 (0.004)		0.002 (0.003)		0.13 (0.18)		0.015 (0.018)	
Child is firstborn		$-0.024^{\dagger}$ (0.014)		0.003 (0.010)		0.65 (0.49)		0.020 (0.052)	
Household head test score (1972)		0.000 (0.004)		−0.007* (0.003)		−0.17 (0.14)		$-0.023^{\dagger}$ (0.013)	
Household head schooling (1972)		−0.004 (0.002)		−0.001 (0.002)		−0.16 (0.11)		$-0.018^{\dagger}$ (0.010)	
Observed "dirty" home		0.000 (0.007)		−0.004 (0.005)		0.59* (0.29)		0.038 (0.028)	
Parental expectations		−0.007 (0.006)		−0.002 (0.005)		−0.16 (0.26)		0.004 (0.023)	
Challenge versus affiliation		0.023 (0.023)		0.027 (0.020)		−0.41 (0.71)		−0.012 (0.078)	
Personal control		0.002 (0.023)		−0.011 (0.015)		0.10 (0.86)		−0.014 (0.081)	
Risk avoidance		−0.004 (0.005)		0.003 (0.002)		−0.12 (0.19)		−0.017 (0.017)	
Birth year dummies?		Yes		Yes		Yes		Yes	
Regression statistics									
R^2 (or pseudo)		.14		.22		.11		.10	
Number of observations		1,137		529		875		875	
p level of test of equality for the three ≤ $25K spline segments		.96		.42		.36		.39	

Note. Sample consists of Panel Study of Income Dynamics children born between 1968 and 1975. Incomes are in 2005 dollars, and childhood incomes are scaled in $10,000. Data in the "different slopes" column show *p* levels of test of equality of within-period < $25K and > $25K slopes. Marginal effects from logistic spline regressions presented for poor health, high distress, and overweight. Poor health regression includes year-of-report dummy variables. Regressions are weighted and standard errors are corrected for multiple children born in the same household. Logistic regression models present marginal effects and standard errors. Marginal effects are interpreted as probabilities and are computed using the MFX command in STATA. BMI = body mass index.

†$p < .10$. *$p < .05$. **$p < .01$.

Table 4

Logistic Spline Regression Models of Childhood Income and Arrests, Jailed, and Nonmarital Births

Childhood income (in $10,000)		Ever arrested (males)			Ever jailed (males)			Nonmarital birth (females)		
				Different slopes			Different slopes			Different slopes
Average annual	< $25K	0.013	(0.055)	*ns*	0.017	(0.028)	*ns*	0.049	(0.046)	*ns*
income (prenatal to age 5)	> $25K	−0.018†	(0.011)		0.001	(0.06)		0.002	(0.012)	
Average annual	< $25K	−0.024	(0.061)	*ns*	−0.004	(0.027)	*ns*	0.013	(0.045)	*ns*
income (ages 6–10)	> $25K	0.016†	(0.009)		−0.001	(0.004)		0.000	(0.009)	
Average annual	< $25K	−0.015	(0.050)	*ns*	0.000	(0.021)	*ns*	−0.035	(0.042)	*ns*
income (ages 11–15)	> $25K	−0.013†	(0.007)		−0.009**	(0.004)		−0.021**	(0.008)	
Other variables										
Black		−0.052†	(0.032)		−0.013	(0.017)		0.135**	(0.049)	
Other minority		−0.039	(.079)		−0.012	(0.044)		0.052	(0.110)	
Child born into intact family		−0.085	(0.054)		−0.041	(0.031)		−0.012	(0.034)	
Child born in South		−0.020	(0.033)		−0.015	(0.018)		−0.048*	(0.024)	
Age of mother at time of birth		−0.005	(0.004)		−0.005*	(0.002)		−0.007*	(0.003)	
Number of siblings		0.020*	(0.010)		0.004	(0.007)		0.020*	(0.009)	
Child is firstborn		0.015	(0.036)		−0.014	(0.016)		−0.006	(0.029)	
Household head test score (1972)		−0.008	(0.008)		−0.002	(0.004)		−0.009	(0.006)	
Household head schooling		0.002	(0.07)		0.004	(0.004)		−0.009†	(0.005)	
Observed "dirty" home		−0.042*	(0.021)		−0.015	(0.010)		−0.015	(0.013)	
Parental expectations		0.030*	(0.015)		−0.003	(0.018)		−0.020†	(0.011)	
Challenge versus affiliation		0.076	(0.050)		0.020	(0.023)		0.044	(0.037)	
Personal control		0.015	(0.045)		−0.002	(0.023)		−0.093*	(0.046)	
Risk avoidance		−0.036**	(0.011)		−0.016†	(0.007)		0.004	(0.009)	
Birth year dummies?		Yes			Yes			Yes		
Regression statistics										
Pseudo R^2		.13			.18			.29		
Number of observations		744			744			778		
p level of test of equality for the three < $25K spline segments		.90			.87			.41		

Note. Sample consists of Panel Study of Income Dynamics children born between 1968 and 1975. Incomes are in 2005 dollars. Childhood incomes are scaled in $10,000. Data in the "different slopes" column show p levels of test of equality of within-period < $25K and > $25K slopes. Regressions are weighted and standard errors are corrected for multiple children born in the same household. Logistic regression models present marginal effects and standard errors. Marginal effects are interpreted as probabilities and are computed using the MFX command in STATA.

$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$.

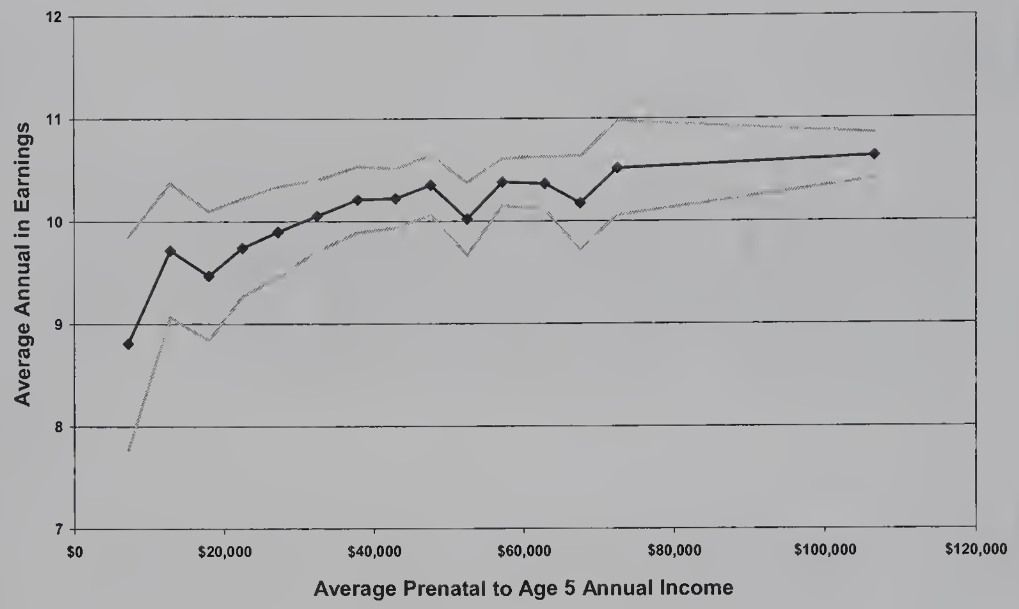


Figure 1. Adult ln earnings by early-childhood income, with 95% confidence intervals.

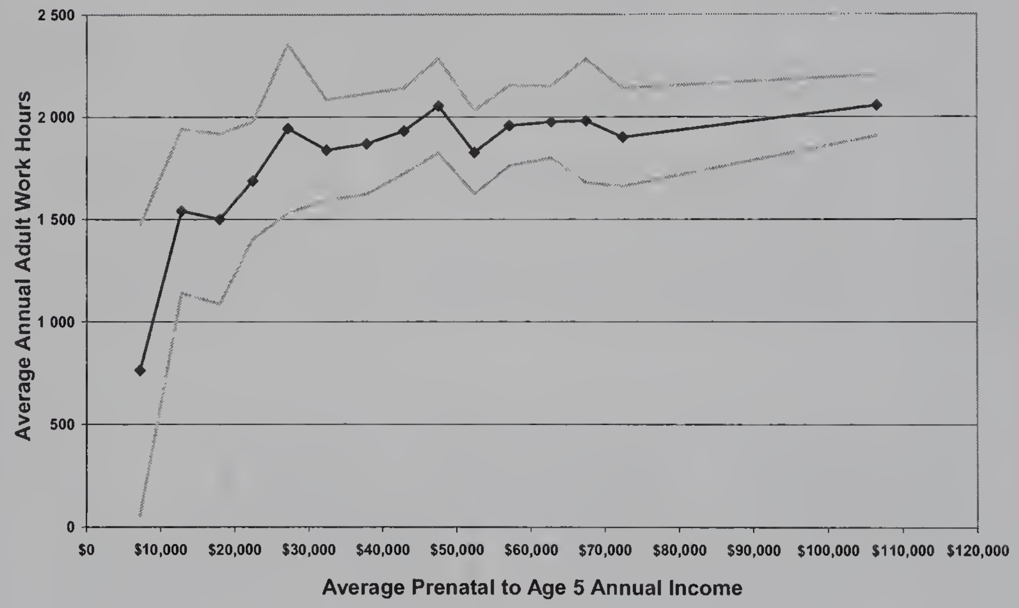


Figure 2. Adult work hours by early-childhood income, with 95% confidence intervals.

somewhat sensitive to outliers, resulting in a decision to truncate both dependent variables at the 99th percentiles of their respective distributions—$4,913 in the case of annual food stamp receipt and $2,954 in the case of annual AFDC/TANF receipt. Additionally, when separate food stamp models for males and females were conducted, we found a much larger and statistically different effect of early poverty for females than for males. The respective coefficients and standard errors for females were −$483 (268) and for males were −$260 (125). The female coefficient has a $p = .06$; the male coefficient has a $p = .04$.

The coefficient on early income in the schooling regression is 0.19 years, which is not statistically significant at conventional levels ($p = .569$). The lack of a significant regression-adjusted association between early income and schooling is at odds with results presented in Duncan et al. (1998). The present analysis includes three different birth cohorts than Duncan et al. and the early income–schooling relation appears somewhat weaker for the new cohorts. A more interesting difference is that Duncan et al. assessed completed schooling by age 20—a kind of "on time" schooling measure. We tested several alternative schooling specifications, including completed schooling by age 21, on-time high school graduation (by age 18 or 19), had dropped out of high school as of age 21, and had attended some college as of age 21. In the case of completed schooling by age 21, we found that the coefficient on early income was 0.338 and statistically significant ($p < .05$). Further, when examining a specification that includes four time periods (prenatal through ages 2, ages 3 through 5, ages 6 through 10, and ages 11 through 15), the coefficient on age 3 through 5 was significant both for completed schooling by age 21 (0.399) and for on-time high school graduation (marginal probability effect from a logistic regression of 0.182). So it appears that early income may matter more for the on-time completion of schooling by the end of adolescent than for the sporadic increases in schooling that often occur later.

Results for the control variables in the completed schooling regression mirror past research, with Black people (adjusted for socioeconomic status and other controls), females, children born first, into intact or smaller families, or born to older mothers, more educated parents, greater household head expectations at birth or greater household head risk avoidance, obtaining more schooling. Few of these controls have persistently significant coefficients across all of the attainment-related outcomes in Table 2.

The marginal (probability change) effects from logistic spline models for poor health, high distress and overweight, and OLS results for BMI are shown in Table 3 and show scattered income effects in middle childhood and adolescence, but not early childhood. In no case are increments to low income in later childhood stages estimated to have a statistically significant impact on any of the health outcomes.

Table 4 presents marginal effects for the behavioral outcomes for men (arrests and jail) and women (nonmarital childbearing). Here again, the early-childhood income segments are not statistically significant in any of these models. As in the case of the health, increments to low income in middle childhood and adolescence are never estimated to have a statistically significant impact on any of the behavioral outcomes.

Extensions to Main Analysis

We explored the robustness of our results in various ways, first by testing for sex and race interactions. We found no statistically significant interactions by sex. For the race interactions, the incarceration of Black men appeared to be significantly less sensitive to increments in early-childhood income than the incarceration of White men. Specifically, the coefficient, expressed as a marginal effect on the probability of occurrence, on White incarceration was 0.38 ($SE = 0.16$) and the Black difference was −0.36 ($SE = 0.16$).

In averaging early-childhood income over the 7-year interval between the prenatal year and fifth birthday, there is a danger that we are missing a narrower sensitive period of income effects. In Table 5, we present results from regressions that are identical to those in Tables 2–4 except that the prenatal to age 5 period is further divided into two segments—the prenatal and birth years and ages 1 through 5. A comparison of coefficients shows that point estimates of income effects for four of the five attainment-related outcomes are all larger when income is measured between ages 1 and 5 than earlier, although only in the case of completed schooling is the difference statistically significant.

The opposite is true of body mass—very early income appears to matter more than income after the birth year, a result explored in some depth in Ziol-Guest, Duncan, and Kalil (2009). This finding of the particular importance of income during the prenatal and birth years for adult BMI may support the ''fetal origins'' hypothesis.

A more technical concern involving the income measure is that it is pretax total family income. An after-tax measure might better capture a family's spending opportunities and constraints. We re-estimated our models using childhood income measures from which PSID-calculated federal income taxes had been subtracted. The results proved somewhat sensitive to outliers but were generally quite similar to those shown earlier.

Table 6 explores the robustness of the prenatal-to-age-5 low-income coefficients in two additional models. The coefficients and standard errors in the first column are identical to those presented in Tables 2–4. In the second column, we retain all of the control variables, but in this case maternal earnings have been subtracted from all three stage-specific income variables. This helps to address the problem that mother's earnings are a component of childhood income. Mothers' labor supply decisions have implications for the amount of time mothers can spend with their children and may be affected by how successful the child's development is viewed by the family. For earnings and annual work hours for which early income effects were statistically significant in Table 2, the new income coefficients are smaller but retain statistical significance.

Stage-specific income results presented in Carneiro and Heckman (2003, p. 120) employ a somewhat different specification than ours in which

Table 5
Coefficients and Standard Errors From Models in Which Prenatal to Age 5 Income Is Divided Into (a) Prenatal and Birth Years and (b) Ages 1–5

	Coefficients and standard errors on low-income < \$25K spline segment		
	Average annual income, prenatal to birth (< 25K)	Average annual income, ages 1–5 (< 25K)	Different slopes
Completed schooling	−0.46 (0.30)	0.63† (0.38)	*p* < .05
ln Earnings	0.17 (0.18)	0.32 (0.21)	*ns*
Annual hours worked	188.70 (131.00)	317.60* (129.23)	*ns*
Annual food stamp receipt	−276.59* (137.75)	−160.59 (133.63)	*ns*
Annual AFDC/TANF receipt (females)	−100.68 (22.58)	−425.68* (220.25)	*ns*
Poor health	0.01 (0.03)	−0.03 (0.03)	*ns*
High distress	0.01 (0.04)	0.00 (0.02)	*ns*
Body mass index	−2.16* (1.08)	−1.45 (2.11)	*ns*
Overweight	−0.11 (0.10)	0.10 (0.11)	*ns*
Arrested (males)	0.02 (0.06)	−0.03 (0.06)	*ns*
Incarcerated (males)	0.02 (0.03)	−0.01 (0.03)	*ns*
Nonmarital childbearing (females)	0.00 (0.05)	0.08† (0.04)	*ns*

Note. Sample consists of Panel Study of Income Dynamics children born between 1968 and 1975. Incomes are in 2005 dollars and childhood incomes are scaled in \$10,000. Coefficient and standard errors from regressions that contain all control variables and four spline segments: prenatal through birth year, ages 1 through 5, ages 6 through 10, and ages 11 through 15. Different slopes indicates whether the low-income segment between prenatal and birth is significantly different from between ages 1 and 5. Ordinary least squares coefficients shown for completed schooling, earnings, annual hours worked, and body mass index; Tobit coefficients shown for food stamps and Aid to Families with Dependent Children/Temporary Assistance to Needy Families (AFDC/TANF); and marginal effects shown for the dichotomous outcomes. Marginal effects are interpreted as probabilities and are computed using the MFX command in STATA.
†$p < .10$. *$p < .05$.

Table 6
Coefficients and Standard Errors on Average Annual Income Prenatal to Age 5 < 25K for Various Model Specifications

	Coefficients and standard error on low-income (< \$25K) spline segment		Controls for permanent (prenatal to age 15) income	
	Basic regression	Exclude maternal earnings from all childhood income measures	Coefficient on low-income permanent income spline segment	Coefficient on low-income prenatal-to-age-5 spline segment
Completed schooling	0.19 (0.33)	−0.25 (0.22)	0.81* (0.36)	−0.11 (0.37)
ln Earnings	0.52* (0.21)	0.36** (0.11)	0.33 (0.20)	0.39† (0.22)
Annual hours worked	506.74** (135.39)	220.65** (78.04)	63.55 (124.30)	446.56** (146.96)
Annual food stamp receipt	−303.14* (152.09)	−102.12 (87.98)	−1152.47** (187.07)	149.98 (195.37)
Annual AFDC/TANF receipt (females only)	−419.18 (264.28)	−90.34 (162.55)	943.17** (340.79)	−570.57 (387.58)
Poor health	−0.01 (0.03)	0.012 (0.015)	0.014 (0.031)	−0.032 (0.0338)
High distress	0.01 (0.012)	0.001 (0.009)	−0.031 (0.030)	0.039 (0.043)
Body mass index	−3.51 (2.45)	−1.57 (1.03)	2.34 (1.47)	−4.62† (2.62)
Overweight	0.03 (0.11)	0.038 (0.063)	0.058 (0.138)	−0.060 (0.140)
Arrested	0.01 (0.05)	−0.004 (0.041)	−0.008 (0.080)	−0.006 (0.071)
Incarcerated	0.02 (0.03)	0.014 (0.020)	0.012 (0.031)	0.006 (0.029)
Nonmarital childbearing	0.05 (0.05)	0.038 (0.029)	−0.082† (0.044)	0.107† (0.058)

Note. Marginal effects reported for the dichotomous outcomes. AFDC/TANF = Aid to Families with Dependent Children/Temporary Assistance to Needy Families.
†$p < .10$. *$p < .05$. **$p < .01$.

childhood income is characterized by permanent (i.e., all-childhood-year average) income and just the early-childhood component. In this case, the coefficient on the early income component shows the coefficient *difference* from permanent income. As shown in the last pair of columns in Table 6, we find that early income is significantly more predictive of earnings and work hours than is permanent childhood income.

One concern in these types of analyses is omitted variable bias, and one approach to deal with this concern is to conduct sibling difference models that capitalize on within-family variation (Johnson & Schoeni, 2007). We tested these models for the earnings and annual work hours regressions. Because of small sample sizes, precision is a problem. For earnings, the coefficient (standard error) on the lower income spline segment was 0.30 (1.08) compared to 0.52 (0.21) in the full sample and the higher income spline was 0.07 (0.09) compared to 0.14 (0.14). For annual work hours, the coefficient and standard error on the lower income spline was 1047 (969) compared to 507 (135) in the full analysis; whereas the higher income early-childhood coefficient in the sibling fixed effects model was −8 (78) compared to 20 (10) in the analysis presented in Table 2.

Given the interesting results for the log earnings models, we sought to explore mediational pathways by sequentially introducing the sets of predictors listed in the columns of Table 7. Controls for childhood maternal work hours and family composition had little impact on the key early-income coefficient. Because controls for work hours and family structure are not so much mediators as additional childhood controls that might be biasing our income estimates, we have not included them in our basic models owing to their potential endogeneity. Completed schooling accounts for very little of the income effect, which is hardly surprising given the weak links between early income and schooling (see Appendix A). The same explanation probably underlies the lack of a mediational role for the behavior and health measures. On the other hand, the inclusion of adult work hours reduces the early-income coefficient by 80%, which suggests that much of the earnings impact of early-childhood income operates through annual work hours rather than wage rate.

The importance of work hours in the link between early-childhood income and annual adult earnings suggests that childhood income should have a relatively small effect on *hourly* earnings in adulthood. When we used average hourly earnings after age 25 as a dependent variable in our standard full-control regression model, the coefficient on early-childhood income was small and statistically insignificant. We investigated the work hours effects further by estimating whether early-childhood income appeared to operate by reducing unemployment or time out of the labor force altogether as opposed to full versus part-time work. For both men and women, the income effects were the strongest for full versus less-than-full time work.

Table 7

Accounting for the Effects of Prenatal to Age 5 Average Annual Income on Log Earnings

	Coefficients and standard error on low-income < $25K spline segment in ln earnings regression					
	Basic model	Add childhood conditions	Add schooling	Add adult behavior	Add adult health	Add adult work hours
Average annual income < $25K (prenatal to age 5)	0.52* (0.21)	0.64** (0.22)	0.60** (0.21)	0.64** (0.21)	0.59** (0.20)	0.10 (0.14)
Average annual income (ages 6–10 and 11–15)	incl.	incl.	incl.	incl.	incl.	incl.
Other background controls	incl.	incl.	incl.	incl.	incl.	incl.
Prenatal to age 15 maternal work hours and family structure		incl.	incl.	incl.	incl.	incl.
Completed schooling			incl.	incl.	incl.	incl.
Arrests, jail, and out-of-wedlock childbearing				incl.	incl.	incl.
Adult health					incl.	incl.
Work hours						incl.
Regression statistics						
R^2	.24	.25	.33	.34	.35	.61
Number of observations	1,016	1,016	1,016	1,016	1,016	1,016

*$p < .05$. **$p < .01$.

Conclusion

Our exploration of the role of economic deprivation early in childhood produced surprisingly strong associations in the case of two important adult attainments—earnings and work hours. These results complement those reported in Duncan and Brooks-Gunn (1997), who found stronger correlations between early-childhood income and childhood achievement than between early income and either childhood behavior or health.

One way of assessing the policy significance of our coefficients is to express our results in income increments that have been associated with past policy changes. With the EITC transferring as much as $4,800 to working families, we used $3,000 as a possible increment. The coefficients imply that a $3,000 annual increase in income between a child's prenatal year and fifth birthday is associated with 19% higher earnings and a 135-hr increase in work hours. Moreover, most of the childhood-income effects on earnings appear to be accounted for by differences in annual work hours, suggesting that the results for work hours are particularly important to understand. Apart from studies looking at adult "idleness" (Haveman & Wolfe, 1994), there is little work on the childhood determinants of full-time work.

One important policy issue is the persistence of these effects across adulthood. A supplemental earnings regression that included an interaction between early-childhood income and age at which adult earnings were measured (not shown in the tables) indicated that the 19% proportionate impact of early-childhood poverty did not change over the ages for which we are able to track the sample (ages 25–37), so we can safely assume that the childhood income impact persists for at least that 13-year period.

We end with the usual cautions regarding the causal interpretation of our income "effects." Among the many approaches to estimating the causal impacts of childhood income on adult outcomes, ours is more sophisticated than some but less sophisticated than others. We are the first study to link high-quality income data across the entire childhood period with adult outcomes measured as late as age 37.

The incomes we observe are determined, in part, by the actions of parents and other family members, which leave our analyses open to omitted-variable bias. Our list of variables controlling for conditions at the time of birth is extensive and ought to reduce a good deal of potential bias. More important, and unusual for studies such as ours, is that our estimates of the impacts of early-childhood income control for income in middle childhood and adolescence. As such, factors such as genetic influences are as likely to affect later and early-childhood income, and thus are controlled, in some degree, by our inclusion of income in other childhood stages. Nevertheless, the possibility of lingering omitted-variable bias remains.

References

Barker, D. (1998). *Mothers, babies, and health in later life.* Edinburgh, UK: Churchill Livingstone.

Barker, D. J. P., Eriksson, J. G., Forsen, T., & Osmond, C. (2002). Fetal origins of adult disease: Strength of effects and biological basis. *International Journal of Epidemiology, 31*, 1235–1239.

Barker, D. J. P., Osmond, C., Forsén, T. J., Kajantie, E., & Eriksson, J. (2005). Trajectories of growth among children who have coronary events as adults. *New England Journal of Medicine, 353*, 1802–1809.

Blau, D. (1999). The effect of income on child achievement. *Review of Economics and Statistics, 81*, 261–276.

Bronfenbrenner, U., & Morris, P. (1998). The ecology of developmental processes. In W. Damon (Series Ed.) & R. M. Lerner (Vol. Ed.), *Handbook of child psychology: Vol. 1. Theoretical models of human development* (5th ed., pp. 993–1028). New York: Wiley.

Carneiro, P., & Heckman, J. (2003) Human capital policy. In J. Heckman & A. Krueger (Eds.), *Inequality in America: What role for human capital policies?* (pp. 77–239). Cambridge, MA: MIT Press.

Centers for Disease Control and Prevention. (2003). *Healthy weight, overweight, and obesity among US adults.* Available at www.cdc.gov/nchs/data/nhanes/databriefs/adultweight.pdf. Accessed on November 1, 2007.

Cunha, F., Heckman, J. J., Lochner, L., & Masterov, D. (2005). Interpreting the evidence on life cycle skill formation. In E. Hanushek & F. Welch (Eds.), *Handbook of the economics of education* (pp. 697–812). Amsterdam: North Holland.

Dahl, G., & Lochner, L. (2005). *The impact of family income on child achievement* (NBER Working Paper No. 11279). Cambridge, MA: National Bureau of Economic Research.

Danese, A., Pariante, C., Caspi, A., Taylor, A., & Poulton, R. (2007). Childhood maltreatment predicts adult inflammation in a life-course study. *Proceedings of the National Academy of Sciences of the United States of America, 104*, 1319–1324.

Duncan, G. J. (1988). The volatility of family income over the life course. In P. B. Baltes, D. L. Featherman, & R. M. Lerner (Eds.), *Life-span development and behavior* (Vol. 9, pp. 317–358). Hillsdale, NJ: Erlbaum.

Duncan, G. J., & Dunifon, R. (1988). Sóft skills and long-run labor market success. In S. W. Polacheck & J. Robst (Eds.), *Research in labor economics* (Vol. 17, pp. 123–150). London: JAI Press.

Duncan, G. J., & Brooks-Gunn, J., (Eds.). (1997). *Consequences of growing up poor.* New York: Russell Sage Foundation.

Duncan, G. J., Brooks-Gunn, J., Yeung, W., & Smith, J. (1998). How much does childhood poverty affect the life chances of children? *American Sociological Review, 63,* 406–423.

Dunifon, R., Duncan, G. J., & Brooks-Gunn, J. (2001). As ye sweep, so shall ye reap. *American Economic Review: Papers and Proceedings, 91,* 150–154.

Fox, N. (1994). The development of emotion regulation: Biological and behavioral considerations. *Monographs of the Society for Research in Child Development, 59*(2/3), 152–166.

Frongillo, E. (2003). Understanding obesity and program participation in the context of poverty and food insecurity. *Journal of Nutrition, 133,* 2117–2118.

Godfrey, K. M., & Barker, D. J. P. (2000). Fetal nutrition and adult disease. *American Journal of Clinical Nutrition, 71,* 1344S–1352S.

Haveman, R., & Wolfe, B. (1994). *Succeeding generations.* New York: Russell Sage Foundation.

Hertzman, C. (1999). The biological embedding of early experiences and its effects on health in adulthood. *Annals of the New York Academy of Sciences, 896,* 85–95.

Holzer, H., Schanzenbach, D., Duncan, G., & Ludwig, J. (2007) *The economic costs of poverty in the United States: Subsequent effects of children growing up poor.* Washington, DC: Center for American Progress.

Johnson, R., & Schoeni, R. (2007). *The influence of early-life events on human capital, health status and labor market outcomes over the life course* (Working paper). Berkeley, CA: Goldman School of Public Policy.

Knudsen, E., Heckman, J., Cameron, J., & Shonkoff, J. (2006). Economic, neurobiological, and behavioral perspectives on building America's future workforce. *Proceedings of the National Academy of Sciences of the United States of America, 103,* 10155–10162.

Mayer, S. (1997). *What money can't buy: The effect of parental income on children's outcomes.* Cambridge, MA: Harvard University Press.

Menard, S. (2004). Six approaches to calculating standardized logistic regression coefficients. *American Statistician, 58,* 218–223.

Poulton, R., & Caspi, A. (2005). Commentary: How does socioeconomic disadvantage during childhood damage health in adulthood? Testing psychosocial pathways. *International Journal of Epidemiology, 34,* 344–345.

Rowe, D., & Rodgers, J. (1997). Poverty and behavior: Are environmental measures nature and nurture? *Developmental Review, 17,* 358. 375.

Sapolsky, R. (2004). Mothering style and methylation. *Nature Neuroscience, 7,* 791–792.

Shonkoff, J., & Phillips, D. (Eds.). (2000). *From neurons to neighborhoods: The science of early childhood development,* Washington, DC: National Academy Press.

Stunkard, A., Faith, M., & Allison, K. (2003). Depression and obesity. *Biological Psychology, 54,* 330–337.

U.S. Census Bureau. (2008). Retrieved February 1, 2008, from http://www.census.gov/hhes/www/poverty/poverty.html

Waddington, C. H. (1957). *The strategy of genes.* Boston: Allen & Unwin.

Waters, E., & Sroufe, L. A. (1983). Social competence as a developmental construct. *Development and Psychopathology, 3,* 79–97.

Yeung, W. J., Linver, M., & Brooks-Gunn, J. (2002). How money matters for young children's development: Parental investment and family processes. *Child Development, 73,* 1861–1879.

Ziol-Guest, K. M., Duncan, G. J., & Kalil, A. (2009). Early childhood poverty and adult body mass index. *American Journal of Public Health, 99,* 527–532.

Appendix A

Weighted Descriptive Statistics of Adult Outcomes by Prenatal to Age 5 Poverty

	Total sample		Income below the official poverty line (A)		Income between 1 and 2 times the poverty line (B)		Income more than twice the poverty line (C)		Significant difference
	M or %	*SD*	*M* or %	*SD*	*M* or %	*SD*	*M* or %	*SD*	
Completed schooling	13.39	2.14	11.83	1.83	12.70	2.02	13.96	2.01	A < B,C; B < C
Unweighted *n*	1,254								
Earnings ($10,000)	34.56	30.93	17.86	16.00	26.79	21.17	39.67	33.96	A < B,C; B < C
Unweighted *n*	1,042								
Annual work hours	1,892.46	699.30	1,512.37	750.25	1,839.12	778.77	1,963.08	641.63	A < B,C; B < C
Unweighted *n*	1,042								
Food stamps	213.75	730.30	896.28	1,453.87	337.28	902.25	70.09	337.43	A > B,C; B > C

Appendix A
Continued

	Total sample		Income below the official poverty line (A)		Income between 1 and 2 times the poverty line (B)		Income more than twice the poverty line (C)		Significant difference
	M or %	*SD*	*M* or %	*SD*	*M* or %	*SD*	*M* or %	*SD*	
Unweighted *n*	1,157								
AFDC/TANF (women only)	158.72	750.60	285.63	804.07	380.61	1,231.28	46.14	344.06	A,B > C
Unweighted *n*	574								
Poor health	0.08	—	0.13	—	0.13	—	0.05	—	A,B > C
Unweighted *n*	1,167								
High distress	0.04	—	0.10	—	0.04	—	0.04	—	*ns*
Unweighted *n*	667								
Body mass index	27.46	5.87	30.63	7.27	28.05	6.70	26.86	5.22	A > B,C; B > C
Unweighted *n*	875								
Overweight	0.61	—	0.79	—	0.61	—	0.59	—	A > B,C
Unweighted *n*	875								
Arrested (men only)	0.16	—	0.26	—	0.21	—	0.13	—	A,B > C
Unweighted *n*	757								
Incarcerated (men only)	0.09	—	0.10	—	0.12	—	0.07	—	B > C
Unweighted *n*	757								
Nonmarital birth (women only)	0.20	—	0.50	—	0.28	—	0.09	—	A > B,C; B > C
Unweighted *n*	778								

Note. In the "significant difference" column, A < B signifies that the mean of those less than the poverty line is statistically significantly smaller than those between 1 and 2 times the poverty line at $p < .05$ (two-tailed). A > B signifies that the mean of those less than the poverty line is statistically significantly larger than those between 1 and 2 times the poverty line at $p < .05$ (two-tailed). Earnings, food stamp value, and Aid to Families with Dependent Children/Temporary Assistance to Needy Families (AFDC/TANF) value in 2005 dollars.

Appendix B

Weighted Descriptive Statistics of Control Variables by Prenatal to Age 5 Poverty

	Total sample		Income below the official poverty line (A)		Income between 1 and 2 times the poverty line (B)		Income more than twice the poverty line (C)		Significant difference
	M or %	*SD*	*M* or %	*SD*	*M* or %	*SD*	*M* or %	*SD*	
Prenatal to age 5, average annual income (2005$)	47,842.48	28,340.64	21,400.11	9,714.14	36,024.01	12,242.22	65,442.38	28,776.41	A < B,C; B < C
Ages 6–10, average annual income (2005$)	54,226.12	39,012.65	23,112.91	16,197.13	41,956.34	23,148.25	73,881.51	42,250.67	A < B,C; B < C
Ages 11–15, average annual income (2005$)	59,067.89	45,369.20	24,743.16	20,140.42	44,137.52	27,222.75	81,649.59	49,337.59	A < B,C; B < C
White	0.77	—	0.38	—	0.70	—	0.88	—	A < B,C; B < C
Black	0.19	—	0.58	—	0.26	—	0.09	—	A > B,C; B > C
Other minority	0.04	—	0.04	—	0.04	—	0.04	—	*ns*
Male	0.53	—	0.52	—	0.52	—	0.54	—	*ns*
Born into intact family	0.84	—	0.44	—	0.78	—	0.94	—	A < B,C; B < C
Born in the South	0.32	—	0.51	—	0.39	—	0.26	—	A > B,C; B > C
Age of mother at birth	24.84	5.76	24.04	7.23	24.13	6.29	25.29	5.15	A,B < C
Number of siblings	2.21	1.79	4.00	3.07	2.56	1.73	1.74	1.16	A > B,C; B > C

Appendix B
Continued

	Total sample		Income below the official poverty line (A)		Income between 1 and 2 times the poverty line (B)		Income more than twice the poverty line (C)		Significant difference
	M or %	*SD*	*M* or %	*SD*	*M* or %	*SD*	*M* or %	*SD*	
Child is firstborn	0.42	—	0.30	—	0.33	—	0.47	—	A, B < C
Household head test score	9.57	2.11	7.71	2.40	9.11	1.99	10.10	1.86	A < B,C; B < C
Household head schooling	12.09	2.94	9.20	3.21	10.78	2.61	13.16	2.39	A < B,C; B < C
Observed "dirty" home	2.18	1.09	2.85	1.09	2.38	1.09	1.80	0.91	A > B,C; B > C
Parental expectations	4.43	1.17	4.09	1.13	4.30	1.20	4.64	1.12	A < B,C; B < C
Challenge versus affiliation	0.66	0.34	0.49	0.33	0.64	0.33	0.74	0.32	A < B,C; B < C
Personal control	0.60	0.35	0.42	0.34	0.55	0.33	0.70	0.33	A < B,C; B < C
Risk avoidance	4.36	1.63	3.34	1.23	3.87	1.45	5.05	1.56	A < B,C; B < C
Unweighted *n*	1,589		297		506		786		

Note. In the "significant difference" column, A < B signifies that the mean of those less than the poverty line is statistically significantly smaller than those between 1 and 2 times the poverty line at $p < .05$ (two-tailed). A > B signifies that the mean of those less than the poverty line is statistically significantly larger than those between 1 and 2 times the poverty line at $p < .05$ (two-tailed).

Child Development, January/February 2010, Volume 81, Number 1, Pages 326–339

From External Regulation to Self-Regulation: Early Parenting Precursors of Young Children's Executive Functioning

Annie Bernier
University of Montreal

Stephanie M. Carlson
University of Minnesota

Natasha Whipple
University of Montreal

In keeping with proposals emphasizing the role of early experience in infant brain development, this study investigated the prospective links between quality of parent–infant interactions and subsequent child executive functioning (EF), including working memory, impulse control, and set shifting. Maternal sensitivity, mind-mindedness and autonomy support were assessed when children were 12 to 15 months old ($N = 80$). Child EF was assessed at 18 and 26 months. All three parenting dimensions were found to relate to child EF. Autonomy support was the strongest predictor of EF at each age, independent of general cognitive ability and maternal education. These findings add to previous results on child stress-response systems in suggesting that parent–child relationships may play an important role in children's developing self-regulatory capacities.

Growing interest has emerged for the hypothesis that early relational experiences are closely related to child neurocognitive development (e.g., Nelson & Bloom, 1997). In fact, the compelling link between quality of parent–child relationships and child cognitive development is often presumed to be due to the interplay between genetic factors and the social environment in shaping early brain development (De Bellis, 2001; Schore, 1996). One aspect of cognitive development that depends heavily on brain maturation is executive functioning (EF). This term refers to the set of higher order cognitive processes that underlie flexible goal-directed behaviors, such as inhibitory control, working memory, planning, and set shifting (Garon, Bryson, & Smith, 2008; Hughes, 2002; Miyake, Friedman, Emerson, Witzki, & Howerter, 2000; Zelazo, Carter, Reznick, & Frye, 1997). While there is convincing evidence that variation within the normal range of EF has important implications for child social and cognitive functioning, research on the antecedents of individual differences in EF lags far behind. Based on the demonstrated experience-dependent maturation of the brain in infancy and on the close connections between EF and brain development (Zelazo, Carlson, & Kesek, 2008), this report examines the role of early parent–child relationships in the development of child EF.

EF in Children

Owing to the close connections between EF and the slow-developing prefrontal cortex (Stuss & Benson, 1986), it was long believed that EF developed in adolescence only (e.g., Golden, 1981). It has since been demonstrated that EF develops much earlier and can be assessed at preschool age (e.g., Carlson, Mandell, & Williams, 2004; Diamond, Barnett, Thomas, & Munro, 2007; Hughes, Dunn, & White, 1998; Zelazo, Müller, Frye, & Marcovitch, 2003) or even toddlerhood (Carlson, 2005; Hughes & Ensor, 2005). In fact, some aspects of EF probably emerge

---

The research described in this article was supported by grants from the Social Sciences and Humanities Research Council of Canada and the Fonds Québécois de Recherche sur la Société et la Culture to the first author. We gratefully acknowledge Marie-Pier Nadeau-Noël, Émilie Rochette, Natasha Ballen, Isabelle Demers, Jessica Laranjo, Célia Matte-Gagné, Stéphanie Bordeleau and several other students for help with data collection, as well as Geneviève A. Mageau, George M. Tarabulsy and Chantal Mongeau for substantial intellectual contributions. Special thanks go to the participating families of the *Grandir Ensemble* project who generously opened their homes to us.

Correspondence concerning this article should be addressed to Annie Bernier, Department of Psychology, University of Montreal, P.O. Box 6128 Downtown Station, Montreal, QC H3C 3J7, Canada. Electronic mail may be sent to annie.bernier@umontreal.ca.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8101-0021

as early as the end of the 1st year of life (Diamond & Goldman-Rakic, 1989), and individual differences assessed in toddlerhood are moderately stable into the preschool years (Carlson et al., 2004; Hughes & Ensor, 2007).

The feasibility of studying EF in children has led to the crucial finding that this set of self-regulatory processes shows meaningful variation within normally developing children of varying ages. One of the most robust findings is that EF is linked to children's theory of mind understanding concurrently, longitudinally, and cross-culturally (e.g., Carlson & Moses, 2001; Carlson et al., 2004; Chasiotis, Kiessling, Hofer, & Campos, 2006; Hughes & Ensor, 2005, 2007; Müller, Zelazo, & Imrisek, 2005; Perner, Lang, & Kloo, 2002; Sabbagh, Xu, Carlson, Moses, & Lee, 2006; Tardif, Wing-Chee So, & Kaciroti, 2007). Another set of findings indicates that EF is related to children's proficiency in mathematics and arithmetic (Blair & Razza, 2007; Bull & Scerif, 2001; Espy et al., 2004; van der Sluis, de Jong, & van der Leij, 2007). In addition, EF has been found to relate to reading ability (Clark, Prior, & Kinsella, 2002; van der Sluis et al., 2007), verbal and nonverbal reasoning (Carlson, Moses, & Breton, 2002; van der Sluis et al., 2007), academic achievement (Biederman et al., 2004; Kinsella et al., 1997), communication, and social skills (Clark et al., 2002), as well as social and moral competence (Kochanska, Murray, & Harlan, 2000; Kochanska, Murray, Jacques, Koenig, & Vandegeest, 1996) and emotion regulation (Carlson & Wang, 2007; Simonds, Kieras, Rueda, & Rothbart, 2007).

Research has thus provided compelling support for the idea that individual differences in EF are meaningful for child cognitive and socioemotional functioning. This stands in sharp contrast to how little we know about the mechanisms that underlie the development of such individual differences. In light of the close connections between EF and frontal brain development (for a review, see Zelazo et al., 2008), and given the increasingly documented impact of environmental input on children's brain development (Chugani et al., 2001; Marshall & Fox, 2004), we propose that the study of early parent–child relationships is likely to be useful in understanding individual differences in children's EF.

Parenting and Child EF

It is increasingly believed that early environmental experiences have a direct impact on brain development (De Bellis, 2001; Gunnar, Fisher, & the Early Experience, Stress and Prevention Network, 2006; Nelson, 2000). The first 2 years of life, especially, witness a spurt in brain growth characterized by overproduction of synapses (Nelson, Thomas, & de Haan, 2006). During this phase, experience is presumed to determine to a large degree which synaptic connections persist and which are selectively eliminated due to lack of use (Greenough & Black, 1992; Nelson & Bloom, 1997; Singer, 1995). General support for such an experience-dependent nature of brain development is provided by animal models (e.g., Gunnar et al., 2006; McEwen, 1999) as well as by the study of grossly inadequate environments, characterized by neglect or abuse (Chugani et al., 2001; Curtis & Cicchetti, 2007; De Bellis, 2001, 2005; Marshall & Fox, 2004; Rutter, O'Connor, & the English and Romanian Adoptee Study Team, 2004; Rutter and the ERA Research Team, 1998). These studies suggest that unfavorable environmental experiences are related to problematic brain development, on both structural (e.g., De Bellis, 2001) and functional levels (e.g., Rutter et al., 2004).

Conversely, it is proposed that favorable environmental experiences, especially those embedded within early caregiving relationships, may have a positive impact on brain development (Glaser, 2000; Nelson & Bloom, 1997; Schore, 1996, 2001). Reviewing studies on maternal care and neural development in animal populations, Francis and Meaney (1999) concluded that, "variations in maternal care that fall within the normal range . . . can still have a profound influence on development. One does not need to appeal to the more extreme conditions of abuse and neglect to see evidence for the importance of parental care" (pp. 132–133). Examining similar questions with human populations, Hane and Fox (2006) found that normative variation in quality of parenting among mother–infant dyads was related to different patterns of frontal electroencephalographic asymmetry in infancy. Hence, given its close connections to brain structures, early development of EF is one aspect of child cognitive development that appears likely to be susceptible to strong care giving influences. According to Glaser (2000), the orderly development of the frontal lobes, strongly implicated in EF, is dependent on "appropriate input and sensitive interaction with the primary caregivers" (p. 101). Given that prefrontal cortex is a slow-developing area, with neural density of the frontal lobes beginning to decline at only about 7 years of age (Huttenlocher, 2002), there is indeed a large window of plasticity during which caregiving can have a significant impact on the developing structures.

Similar claims are made at the functional level. For instance, Kopp (1982) suggested that early caregiving is a central mechanism in children's developing capacities for self-regulation, the hallmark of EF. In line with this, attachment researchers have often proposed that caregivers initially act as external regulators of the infant's rhythms and affect, gradually facilitating the child's increasing capacity to self-regulate (e.g., Grossmann & Grossmann, 1991; Hofer, 1995; Spangler, Schieche, Ilg, Maier, & Ackermann, 1994). This view is supported by several studies that found links between quality of care and child psychophysiological regulation (see Gunnar & Donzella, 2002, for a review). Parenting may thus be a crucial factor in the development of child EF. However, it may not be optimal to consider parenting as a unitary construct, as empirical research with both animal and human populations increasingly suggests that different dimensions of caregiving can have distinct contributions to child functioning (e.g., Hofer, 1996; Meins, Fernyhough, Fradley, & Tuckey, 2001; Moran, Forbes, Evans, Tarabulsy, & Madigan, 2008).

Carlson (2003) proposes three dimensions of parenting likely to favor the development of child EF: maternal sensitivity, scaffolding, and mind-mindedness. Sensitivity, which consists of appropriate and consistent responses to infants' signals, would provide them with successful experiences of impacting the social environment. Scaffolding, that is, offering children age-appropriate problem-solving strategies, is likely to yield successful experiences of problem-based learning. Finally, mind-mindedness, or the parent's tendency to use mental terms while talking to the child, is thought to offer children verbal tools with which to progress from being externally regulated to self-regulated. Hence, these three parental behaviors represent ways in which caregivers act as external regulators, which according to Harrist and Waugh (2002), serves to help children become gradually more self-regulating away from caregivers. However, each behavior also has distinctive features, and hence possibly distinct contributions to child EF. Sensitivity and scaffolding refer to what the mother does, while mind-mindedness refers to what she says to the child about inner states, and sensitivity and scaffolding are likely to be activated in distinct situations, that is, contexts of personal distress versus exploration, respectively.

The Present Study

Although animal studies have documented the impact of parental care on the development of prefrontal systems (Gunnar et al., 2006), human research is limited, pertaining mostly to highly inadequate caregiving. Accordingly, the main purpose of this report was to examine the prospective links between quality of early mother–infant interactions and child subsequent EF in a normative population. The measurement of mother–infant interactions proceeded along the three dimensions proposed by Carlson (2003): We assessed maternal sensitivity and mind-mindedness at 12 months and maternal autonomy-support, including scaffolding, at 15 months. Child EF was assessed at 18 and 26 months. It was expected that more favorable parenting, indicated by higher sensitivity, more mind-minded comments, and greater autonomy support during mother–infant interactions, would be associated with better child performance on EF tasks at both 18 and 26 months of age, above and beyond parental education and child cognitive functioning. A secondary aim was to continue the recent line of work that has allowed for reliable measurement of EF at increasingly younger ages (Carlson, 2005; Hughes & Ensor, 2005). Hence, while the timing of the second assessment was based on the fact that validated measures of EF start at 2 years of age, the first assessment was planned 6 months before this age to examine the feasibility of assessing earlier manifestations of EF. We thus developed EF tasks for 18-month-olds and examined their capacity to detect individual differences, as well as continuity between child performance at 18 and 26 months of age.

Method

Participants

Eighty middle-class mother–infant dyads (44 girls and 36 boys) living in a large Canadian metropolitan area participated in this study. Families were recruited from birth lists provided by the Ministry of Health and Social Services. Criteria for participation were full-term pregnancy and the absence of any known physical or mental disability or severe developmental delay in the infant. Family income varied from less than $20,000 CDN to over $100,000 CDN, with an average of $70,000 CDN. Mothers were between 19 and 44 years old ($M = 29.4$). Most (78.8%) had a college degree and 81.3% were Caucasian. All but seven were married or living with the child's father throughout data collection.

Procedure

The dyads took part in four visits, when the child was 12–13 (T1; $M = 12.9$), 15 (T2; $M = 15.5$), 18 (T3; $M = 18.3$), and 26 (T4; $M = 26.3$) months of age. Home visits were conducted at T1, T2, and T4, while T3 consisted of a laboratory visit. All visits lasted between 70 and 90 min. The first home visit was modeled after the work of Pederson and Moran (1995), and aimed at challenging the mother's capacity to divide her attention between several competing demands, thus reproducing the natural conditions of daily life when caring for an infant. Research tasks included a developmental assessment (the Bayley scales), a brief interview with the mother, a 10-min videotaped mother–infant free play, and questionnaires that mothers had to complete while the infant was *not* looked after or kept busy by the research assistants. Maternal sensitivity was assessed with the Maternal Behavior Q-Sort (MBQS; see below) based on observations made throughout this visit, and mind-mindedness was later coded from the videotaped play interaction. The second home visit was similar in length and structure to the first one. Mothers were asked to help their children complete two puzzles that were designed to be slightly too difficult for the infants, such that they would require some adult assistance to complete them. This interaction was videotaped and later coded for maternal autonomy-supportive behaviors. Other activities included an interview with the mother, free play, and questionnaires. The third visit took place in the laboratory. The EF tasks described below were administered, along with other child and dyadic activities not used in the present study. Finally, the fourth visit took place in the families' homes. Most of the visit consisted of the administration of the EF tasks described below, along with other child and dyadic activities not included in this report.

Measures

Parenting

Maternal sensitivity. Maternal sensitivity was assessed using the MBQS (Pederson & Moran, 1995), a 90-item measure designed to assess the quality of maternal behaviors during in-home mother–infant interactions. The observer noted maternal behaviors throughout the visit and rated the MBQS immediately afterward based on the entire observation period. Items describing potential maternal behaviors are sorted into nine clusters, ranging from very similar to very unlike the observed mother's behaviors. The observer's sort is then correlated with a criterion sort representing the prototypically sensitive mother, which is provided by the developers of the instrument. Sensitivity scores thus vary from −1 (*least sensitive*) to 1 (*prototypically sensitive*).

The development of the MBQS is anchored in attachment theory and specifically in the descriptions of sensitive responsiveness provided by Ainsworth, Bell, and Stayton (1974). Pederson and colleagues (Pederson, Gleason, Moran, & Bento, 1998; Pederson & Moran, 1995; Pederson et al., 1990) have presented detailed descriptions regarding the development of the MBQS, as well as its validity and reliability. These authors' longitudinal studies show that the MBQS is useful in predicting multiple aspects of infant development. Moreover, the MBQS is significantly correlated with other assessments of maternal behavior, such as the Home Observation for Measurement of the Environment Inventory and the Ainsworth scales (see Pederson & Moran, 1995).

Previously trained home visitors performed their first few home visits with a more experienced colleague, and these visits were followed by a debriefing session aimed at reviewing the salient elements of the visit before scoring the MBQS. Following this training period, 25 home visits (31%) were conducted by two research assistants, who completed the MBQS independently. Agreement between the two raters' sorts was high, intraclass correlation (ICC) = .89.

Maternal mind-mindedness was assessed through a 10-min free-play sequence between mother and infant, in which the mother was instructed to play as she normally would, but with toys that were brought by the research team. Videotaped interactions were later coded by a trained assistant using Meins et al.'s (2001) coding system. Five categories of comments were assessed: (a) on the infant's mental state, such as thoughts, desires, knowledge (e.g., "You want this book"); (b) on mental processes (e.g., "You find this game difficult"); (c) on the infant's emotional engagement (e.g., "You had enough"); (d) on the infant's attempts to manipulate other people's thoughts (e.g., "You're making fun of me"); and (e) comments that involved the mother speaking for the infant (e.g., "See mom, it's easier this way''). Each comment was then coded as appropriate or inappropriate according to Meins et al.'s guidelines. The number of appropriate comments in each category were summed into a total score, which is used in all subsequent analyses.

A randomly selected 42.5% (*n* = 34) of videotapes were independently coded by a second trained rater. Interrater reliability was satisfactory, ICC = .89. Mind-mindedness has good convergent and predictive validity with measures of maternal sensitivity, mother–child attachment security, and child theory of mind understanding (Meins et al., 2001; Meins et al., 2002).

Maternal autonomy support. The notion of autonomy-support stems from Self-Determination Theory and refers to parental behaviors aimed at supporting children's goals, choices, and sense of volition (Grolnick & Ryan, 1989). Mother–infant dyads were asked to complete a challenging problem-solving task (two puzzles) together during a second home visit when the child was 15 months of age. Following the Grolnick, Frodi, and Bridges (1984) rating system, maternal behaviors were coded along two scales: verbal and nonverbal behaviors. In the current study, we further divided these scales into four categories reflecting the specific behaviors that constitute autonomy-support (Grolnick, Gurland, DeCourcey, & Jacob, 2002). The four Likert scales assess the extent (1–5) to which the mother (a) intervenes according to the infant's needs and adapts the task to create an optimal challenge (equivalent to scaffolding); (b) encourages her child in the pursuit of the task, gives useful hints and suggestions, and uses a tone of voice that communicates to the child that she is there to help; (c) takes her child's perspective and demonstrates flexibility in her attempts to keep the child on task; and (d) follows her child's pace, provides the child with the opportunity to make choices, and ensures that the child plays an active role in the completion of the task. Given the intercorrelations among the four scales (ranging from .51 to .82), they were averaged into a total autonomy support score (α = .89). All videotapes were coded by the third author of this report, and 38 of the 80 (47.5%) interactions were independently coded by a second observer. ICC between coders for the total autonomy-support score showed excellent interrater agreement, ICC = .86.

Child EF: 18-Month Assessment

The earliest measurement guidelines currently available to assess EF start at 24 months (Carlson, 2005). However, neuropsychological testing suggests that certain aspects of executive functions begin to develop earlier, notably rule use and working memory. We thus developed two downward adaptations of EF tasks originally designed for 2-year-old children.

Hide the Pots (adapted from Hughes & Ensor, 2005 [Spin the Pots]; see also Uzgiris & Hunt, 1975 [Scale I, Search for Hidden Objects]). An attractive sticker was hidden in full sight of the child under one of three opaque pots of different colors. In the warm-up phase, no further action took place, and the child was immediately asked to retrieve the sticker where it was hidden. Three practice trials were conducted (one with each pot). This phase was intended to introduce children to what was expected. In the second phase, the experimenter also hid the sticker once under each pot, but the pots were covered with a blanket before the child was invited to find the sticker on each trial. The child thus had to hold the location of the sticker in memory, remove the blanket, and reach for the pot where the sticker was hidden. Scores consisted of the number of trials (0–3) where the child found the sticker on his or her first attempt.

Categorization (adapted from Carlson et al., 2004). We used only the first phase of the original reverse categorization task, because the reverse phase was found to be difficult even for older toddlers (Carlson, 2005). Children were taught a categorization rule whereby baby animals were sorted together in the ''baby box,'' whereas adult animals were placed into the ''mommy box.'' The experimenter used the first six animals (three adults and three babies) to show the child how to sort the toys, and then asked the child to sort the six remaining ones. As many children placed all six animals together in one box (and thus got three ''correct'' answers by accident), the score was computed as the number of dyads (e.g., mother and baby zebra) that were correctly sorted separately by the child (0–3).

Child EF: 26-Month Assessment

The tasks conducted at 26 months were chosen based on Carlson's (2005) measurement guidelines with the aim of maximizing detection of individual differences in three dimensions of EF: working memory, inhibitory control, and set shifting.

Spin the Pots (Hughes & Ensor, 2005). Children were asked to search for stickers that were hidden in opaque pots, subsequently covered and rotated. Six stickers and eight pots of very different visual appearances were used. Each time a sticker was found, the pots were covered and rotated again. The score was calculated as 16 minus the number of errors made (i.e., looking under a pot in which no sticker had been hidden or perseverative looking).

Delay of Gratification (Kochanska et al., 2000). The experimenter placed a present under a transparent cup and asked children to wait until she rang a bell before retrieving it. Four trials were conducted, where the child had to wait 5, 10, 15, and then 20 s. Scores were the number of seconds waited on each trial.

Shape Stroop (Kochanska et al., 2000). Children were first shown six cards depicting three small and three large fruits, and were asked to point to each in turn to ensure they knew the names and sizes of the fruits (e.g., ''Show me the big apple''). They were then shown three cards each depicting one of the small fruits embedded in one of the larger ones, and asked to point to each of the small fruits in turn (e.g., ''Show me the *small* banana''). The score consisted of the number (0–3) of small fruits correctly pointed to.

Baby Stroop (adapted from Hughes & Ensor, 2005). Children learned a rule for feeding two dolls, feeding the ''mommy'' doll with a larger spoon and the baby doll with a smaller spoon. As soon as the child clearly understood the rule, it was reversed such that the larger doll had to be fed with the smaller spoon, and vice versa. Scores thus ranged from 0 to 2.

At both T3 and T4, administration of the EF tasks was videotaped for later coding. Because we were interested in individual differences rather than group patterns, EF tasks were administered in fixed order.

Child Cognitive Ability

In order to tap into what is unique to EF beyond general cognitive functioning, we used the Mental Development Index of the Bayley Scales of Infant Development (MDI; Bayley, 1993) at 12 months. The MDI is a well-standardized test of cognitive development for children aged 1–42 months, which shows predictive validity with respect to later verbal and cognitive performance (e.g., Lemelin, Tarabulsy, & Provost, 2006). A total of 23 protocols (29%) were double-coded by independent raters (using the video recording of the administration). ICC between raters' MDI scores suggested good reliability, ICC = .85.

Results

Preliminary Analyses

Table 1 presents descriptive statistics for all parenting and EF variables. The three parenting variables showed excellent variability (note that sample size varies slightly due to technical difficulties with certain videotapes). EF tasks also showed adequate

Table 1
Mean, Standard Deviation, and Range for All Parenting and Executive Functioning (EF) Measures

Variable	*N*	Mean	Standard deviation	Observed range	Theoretical range
Parenting					
Maternal sensitivity	80	0.59	0.33	−0.60 to +0.86	−1.0 to +1.0
Maternal mind-mindedness	74	14.58	8.71	0–39	0–N/A
Maternal autonomy support	77	3.28	1.13	1–5	1–5
Child EF 18 months					
Hide the Pots	80	1.22	0.91	0–3	0–3
Categorization	80	0.61	1.17	0–2	0–3
Child EF 26 months					
5-s delay	80	4.67	0.95	0–5	0–5
10-s delay	80	8.99	2.36	0–10	0–10
15-s delay	80	11.64	5.17	1–15	0–15
20-s delay	80	14.08	7.78	1–20	0–20
Spin the Pots	80	10.45	3.41	4–16	4–16
Shape Stroop	80	1.50	1.19	0–3	0–3
Baby Stroop	80	0.37	0.66	0–2	0–2

variability, although Categorization at 18 months and Baby Stroop at 26 months were negatively skewed, whereas the 5- and 10-s Delay of Gratification trials were positively skewed.

The two EF tasks at 18 months of age were unrelated to one another, $r = .11$, *ns*. These tasks were thus considered separately in all subsequent analyses. Table 2 shows the intercorrelations between EF scores at 26 months. In line with previous studies of EF in toddlers (e.g., Carlson et al., 2004; Hughes & Ensor, 2005), scores were mildly to moderately correlated. They were submitted to a principal component analysis, which yielded a two-factor solution (Eigen values > 1.0) representing 64.7% of the total variance. These two factors were submitted to a principal axis rotation (oblimin). Factor loadings for 10-s delay (.85), 5-s delay (.83), 15-s delay (.79), and 20-s delay (.71) suggest that the first factor taps impulse control, whereas the second factor appears to represent working memory, set shifting, and inhibitory control (Conflict EF): Spin the Pots (.82), Baby Stroop (.67), and Shape Stroop (.59). No cross-loadings (above .25) were observed, and the correlation between the two factors was .32. Given that the factor structure was very clear empirically and reproduced the two dimensions of Impulse Control and Conflict previously identified with young children (Carlson & Moses, 2001; Carlson et al., 2004), two averaged standardized scores were computed and used in further analyses. Impulse Control and Conflict were moderately correlated, $r = .28$, $p < .05$.

We next examined whether sociodemographic variables (child gender, number of siblings, maternal and paternal age and education, and family income) or child cognitive ability were related to EF performance at 18 and 26 months of age. Two relations reached significance at 18 months: Children of more educated mothers ($r = .22$, $p < .05$) and children who had higher mental development scores on the Bayley ($r = .24$, $p < .05$) performed better on Hide the Pots. Maternal education and child overall cognitive functioning (Bayley scores; $r = .17$, $p < .10$) were thus used as covariates in all further analyses pertaining to EF at 18 months. The analyses revealed no significant associations between the two EF dimensions at 26 months and sociodemographics (all *r*s < .18, *p*s > .10). However, child cognitive functioning was positively related to both EF dimensions: Impulse Control ($r = .28$, $p < .05$) and Conflict EF ($r = .25$, $p < .05$). We thus covaried cognitive functioning in all subsequent analyses involving EF at 26 months. Age was unrelated to EF performance, which was expected given the restricted age range at each time point. Child age was therefore not retained as a covariate in the main analyses.

Table 2
Intercorrelations Among Executive Functioning Scores at 26 Months

	10 s	15 s	20 s	Spin the Pots	Shape Stroop	Baby Stroop
5 s	.59***	.51***	.28*	.06	.22†	.10
10 s	1.00	.60***	.50***	.03	.19	.09
15 s		1.00	.63***	.11	.21†	.10
20 s			1.00	.24*	.28*	.12
Spin the Pots				1.00	.33**	.22†
Shape Stroop					1.00	.19
Baby Stroop						1.00

Note. 5 s = 5-s delay; 10 s = 10-s delay; 15 s = 15-s delay; 20 s = 20-s delay.
$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

In line with their definitions as distinct aspects of one global construct, the three parenting dimensions were mildly intercorrelated: sensitivity and mind-mindedness ($r = .36$, $p < .01$), sensitivity and autonomy support ($r = .26$, $p < .05$), and mind-mindedness and autonomy support ($r = .38$, $p < .01$). These three dimensions were thus retained as separate scores.

Main Analyses

We first examined whether our adaptations of tasks for 18-month-olds tapped into early manifestations of EF by examining longitudinal stability at 18 and 26 months of age. The results showed that Hide the Pots at 18 months was related to both EF dimensions at 26 months (Impulse Control: $r = .28$, $p < .05$; Conflict: $r = .25$, $p < .05$), while Categorization was associated with later performance on Conflict only, $r = .24$, $p < .05$.

Table 3 presents the bivariate correlations between the three parenting dimensions and child performance on the two EF tasks at 18 months, as well as the partial correlations when accounting for maternal education and child cognitive functioning. Results indicated that child performance on Hide the Pots was related to all three parenting dimensions, but marginally so in the case of maternal sensitivity. Categorization was related to autonomy support. As expected, relations were positive in all cases.

We next examined relations between the parenting dimensions and child performance on the two EF dimensions at 26 months. Table 4 presents

Table 3

Bivariate (and Partial) Correlations Between the Three Parenting Dimensions and Child Performance on the Two EF Tasks at 18 months

EF task	Maternal sensitivity	Maternal mind-mindedness	Maternal autonomy support
Hide the Pots	.20† (.18)	.35** (.33**)	.38** (.39**)
Categorization	.08 (.15)	.04 (.04)	.25* (.30*)

Note. Numbers vary between 74 (mind-mindedness) and 80 (sensitivity). Partial correlations while controlling for maternal education and child cognitive functioning are shown in parentheses. EF = executive functioning.
†$p < .10$. *$p < .05$. **$p < .01$.

Table 4

Bivariate (and Partial) Correlations Between the Three Parenting Dimensions and Child Performance on the Two EF Dimensions at 26 Months

EF dimension	Maternal sensitivity	Maternal mind-mindedness	Maternal autonomy support
Impulse Control	.07 (.03)	.22† (.19)	.13 (.10)
Conflict	.25* (.26*)	.23† (.18)	.31* (.32*)

Note. Numbers vary between 74 (mind-mindedness) and 80 (sensitivity). Partial correlations while controlling for child cognitive functioning are shown in parentheses. EF = executive functioning.
†$p < .10$. *$p < .05$.

the bivariate correlations as well as the partial correlations when holding child cognitive functioning constant. All three parenting dimensions predicted later Conflict EF (although the relation with mind-mindedness was marginally significant only). Impulse Control was marginally related to maternal mind-mindedness, but this relation became nonsignificant when accounting for child cognitive functioning.

In summary, autonomy support was found to relate to three of four EF measures, that is, both 18-month tasks and Conflict at 26 months. In all cases, these relations held when controlling for potential confounds that were measured. Maternal sensitivity was associated with Conflict at 26 months and tended to relate to Hide the Pots at 18 months, although this latter relation became nonsignificant when controlling for potential confounds. Finally, mind-mindedness was significantly related to Hide the Pots at 18 months and marginally associated with both EF dimensions at 26 months, but these two latter associations did not hold over and above shared variance due to child cognitive functioning.

Table 5

Summary of Regression Analyses Predicting Child Performance on Hide the Pots at 18 Months

Block	R^2	ΔR^2	*F* change	β
1. Maternal education				.16
Child cognitive functioning	6.4%	6.4%	1.78*	.18
2. Maternal education				.04
Child cognitive functioning				.10
Autonomy support				.35*
Mind-mindedness	23.9%	17.5%	5.77**	.15

Note. $N = 74$.
*$p < .05$. **$p < .01$.

Note that in all cases where either maternal sensitivity or mind-mindedness related significantly to child EF, autonomy support related to the same EF index. We thus ran regression analyses to decipher the common and unique contributions of each parenting dimension to the variance in indices of EF that related to more than one parenting dimension (i.e., Hide the Pots at 18 months and Conflict EF at 26 months). Only those control and parenting variables that were significantly related to the EF index of interest were entered in each regression equation. Table 5 presents the results of the first hierarchical regression analysis predicting child performance on Hide the Pots. Maternal education and child cognitive functioning were entered in a first block, followed by mind-mindedness and autonomy support in a second block. Table 5 shows that the model predicted 23.9% of the variance in Hide the Pots, $F(4, 74) = 3.65$, $p < .01$, with parenting contributing a unique 17.5% ($p < .01$) over and above what is explained by maternal education and child cognitive functioning. The regression coefficients indicate that while autonomy support was uniquely related to child performance on Hide the Pots when mind-mindedness was accounted for ($\beta = .35$, $p < .05$), mind-mindedness did not relate to Hide the Pots when autonomy support was accounted for ($\beta = .15$, *ns*). Note that this pattern of results is consistent with statistical mediation; we return to this in the Discussion.

Table 6 presents the results of the second hierarchical regression analysis with Conflict EF at 26 months as the dependent variable. Child cognitive functioning was entered in a first block, followed by maternal sensitivity and autonomy support in a second block. The overall model was significant, $F(3, 77) = 2.89$, $p < .05$, explaining 16.4% of the variance of child performance on Conflict. A unique 10.1% ($p < .05$) was explained by parenting above and beyond child cognitive functioning.

Table 6

Summary of Regression Analyses Predicting Child Performance on Conflict EF at 26 Months

Block	R^2	ΔR^2	F change	β
1. Child cognitive functioning	6.3%	6.3%	3.11*	.25*
2. Child cognitive functioning				.19
Autonomy support				.25*
Sensitivity	16.4%	10.1%	3.59*	.14

Note. $N = 77$. EF = executive functioning.
*$p < .05$.

Quite similar to what was found with Hide the Pots, only autonomy support remained uniquely related to Conflict in the final model (β = .25, $p < .05$).

In a last step of this analysis, we examined whether parenting predicted EF at 26 months over and above stability in EF between 18 and 26 months. Given its clearer connections to EF at 26 months, we used Hide the Pots as the baseline 18-month EF index. Two regression analyses were conducted, one on Conflict as the dependent variable and one on Impulse Control. Each equation consisted of 18-month Hide the Pots and child cognitive functioning in a first block, followed by all three parenting dimensions in a second block. Analyses revealed that parenting did not contribute to changes in Impulse Control (4.3%, *ns*) but marginally predicted Conflict after accounting for longitudinal stability between 18 and 26 months and child cognitive functioning, 12.4% ($p < .10$). Contrary to expectations, however, this trend was not attributable to autonomy support (β = .16, *ns*) or sensitivity (β = .10, *ns*), but rather to mind-mindedness (β = .32, $p < .05$). Hence, while autonomy support was related to EF at both 18 and 26 months, mind-mindedness better explained increases in child performance between these two ages.

Discussion

The main purpose of this study was to investigate the prospective links between quality of parent–infant interactions and subsequent child EF. It was expected that higher quality parenting would predict better child EF performance at both short-term (3–6 months) and long-term (14 months) follow-ups. The results lend general support to this hypothesis. We found that mothers who were more sensitive with their 12-month-old child had children performing better on Conflict EF at 26 months. Their children also tended to perform better on working memory at 18 months, but this trend did not hold above the controls. Children whose mothers were more mind-minded when they were 12 months old performed better on working memory at 18 months, and tended to perform better on both Conflict and Impulse Control at 26 months, although these two latter trends did not hold when controlling for child cognitive functioning. However, mind-mindedness was related to changes in child performance between 18 and 26 months. Finally, mothers who were more autonomy supportive when their child was 15 months of age had children performing better on working memory and categorization at 18 months as well as on Conflict EF at 26 months. These latter findings held up over the controls.

These results provide support for the assumption that caregiving is involved in children's developing self-regulatory abilities (Hofer, 1995; Kopp, 1982) and suggest that the influence of parenting on the development of children's self-regulation can be observed not only on psychophysiological dimensions (Gunnar & Donzella, 2002) but also in the behavioral and cognitive domains. It remains to be seen whether the results found here suggest that early caregiving impacts brain structures involved in EF or indicate that parenting provides the child with the social context in which to practice emerging regulatory skills. These two possibilities, however, are not mutually exclusive and in fact, it is likely that a transactional process involving both is at play. It has been argued (e.g., Kraemer, 1992) that the effects of early caregiving on brain development, observed in animal studies and samples of maltreated children, may be observable also with normally developing infants. Such an influence could be attributable to the impact of parenting on child stress-response systems (Gunnar et al., 2006). The ensuing orderly development of frontal brain areas would favor the emergence of executive functions, allowing the child to regulate behavior more effectively at an early age. This emerging self-regulation may contribute to the maintenance of harmonious parent–infant interactions, in turn providing the child with a context favorable to increased self-regulation and possibly contributing further to brain development. Hence, early parenting may impact the process both by affecting the child's developing neurobiological structures and through the social environment provided to the child, and bidirectional effects between parenting and child EF are likely.

This is among the first studies to document links between early parent–child relationships and later child EF. In fact, extremely little is known about any antecedents of child EF, besides intrinsic factors such as child cognitive functioning (e.g., Carlson et al., 2004; Hughes & Ensor, 2005) and more distal factors such as socioeconomic status (SES; e.g., Mezzacappa, 2004). Hughes and Ensor (2005) examined relational factors and found that although quality of parenting was positively related to child EF in 2-year-olds, this relation did not hold after accounting for child age and verbal ability. The difference in results between their study and ours is particularly intriguing given that in both studies, great care went into the assessment of parenting, using both live observation and videotaped interactions in a variety of contexts, and multi faceted coding systems. In addition, our prospective design constituted a stringent test, making shared method variance especially unlikely to inflate the results. One explanation for the discrepant findings may relate to the nature of the samples: While our sample was predominantly middle-class, most families in Hughes and Ensor's sample were from lower SES backgrounds. Growing evidence suggests that the impact of caregiving on child outcomes may vary according to the social conditions of the dyad (see, e.g., Bernier and Meins's, 2008, analysis of parenting and attachment in low- vs. high-risk samples). One may speculate that in certain low-SES families, factors such as family chaos or parenting stress may limit the opportunities for joint cognitive activity, while the presumably lower stress experienced by middle-class families may leave more room for competent caregiving to foster development of the child's neural structures and self-regulatory capacities.

The clearest finding from this study is that autonomy support is the aspect of parenting that was most robustly related to age-specific indices of child EF. Autonomy support consists mainly of scaffolding, respecting the child's rhythm, and ensuring that he or she plays an active role in successful completion of the task. One can easily see how this set of behaviors is likely to provide the child with a sense of accomplishment and self-efficacy (cf. Bandura, 1997). That autonomy support would favor the emergence of child EF thus appears sensible, and we would argue that this is likely to be a robust finding. As well, it is possible that the autonomy support measure, which involved externally guided problem solving, provided a more direct contribution than other parenting qualities to self-guided problem solving that was required on the EF tasks administered later. In fact, our results complement the large literature on scaffolding, a key aspect of autonomy support, which has often been found to impact other aspects of child cognitive development such as language and problem-solving competence (e.g., Landry, Garner, Swank, & Baldwin, 1996; Landry, Smith, Swank, & Miller-Loncar, 2000).

In terms of the respective contributions of the different indices of parenting, however, one needs to take into account the methodological parameters of this study before drawing firm conclusions. Our results would appear to suggest that autonomy support not only relates more consistently to child EF than maternal sensitivity and mind-mindedness but also that it is responsible for most links found between child EF and either sensitivity or mind-mindedness. Recall that the regression analyses revealed that when two parenting dimensions were entered concurrently to predict an age-specific aspect of child EF, we did not find additive contributions, but rather that autonomy support remained the only significant predictor when either sensitivity or mind-mindedness was accounted for. A possible explanation is that this pattern of results, consistent with a statistical mediation, does indeed reflect true mediation. It stands to reason that a mother who is skilled at interpreting her child's signals correctly and responding appropriately (sensitivity), or is well aware of her child's mental processes while interacting with him or her (mind-mindedness), is well equipped to provide the child with an autonomy-supportive environment characterized by appropriate scaffolding, perspective taking, and respect for the child's rhythm in problem solving. However, the design of the current study does not permit us to rule out an alternative, methodological explanation. Autonomy support was assessed closer in time to EF than sensitivity and mind-mindedness (at 15 months vs. 12 months). Thus, the fact that autonomy support appeared to mediate the relations between sensitivity or mind-mindedness and child EF might also be attributable to different degrees of temporal proximity.

When predicting increments in EF performance, we found a slightly different picture: Mind-mindedness remained the only significant predictor of 26-month Conflict EF after accounting for child cognitive functioning, 18-month EF, and the other parenting indices. Combined with the finding that autonomy-support is related to both 18- and 26-month EF performance, this suggests that while autonomy support accounts for stability in EF performance, mind-mindedness rather accounts for changes between these two age periods. However, given that regression coefficients can be quite sensitive to the inclusion of correlated predictors in the equation, one should probably interpret the findings cautiously. While these analyses reaffirm the role of parenting in the *development* of EF through an especially stringent test, they also highlight the need for replication in larger samples before drawing firm conclusions on the role of specific aspects of parenting in changes and stability in EF. Similarly, given that part of what is observed with the Bayley scales may be presumed to be executive in nature, one may speculate that the analyses controlling for the MDI may actually be underestimating the predictions from parenting variables.

Nonetheless, the findings with mind-mindedness appear to suggest that parent attunement and talk about the child's mental states can build on basic levels of EF that were created, in part, by autonomy-supportive parental behavior, and give the child verbal tools to expand further on emerging regulatory abilities. This is speculative however, and teasing apart the true contributions of distinct aspects of caregiving to the development of child EF will necessitate assessment of different parenting indicators under methodologically equivalent conditions (including delay with subsequent assessment of EF). As well, research with higher risk samples may yield more variation in the lower spectrum of all three caregiving indicators, thus providing more power to detect links to child EF.

An important connection to previous research concerns the concept of effortful control (Blair & Razza, 2007; Kochanska et al., 1996; Kochanska et al., 2000; Rueda, Posner, & Rothbart, 2005). Whereas EF has its origins in neuropsychological research, effortful control refers to temperamentally based differences in the control of attention (Rothbart, 1989). Based on previous work suggesting that Kochanska's tasks provided excellent measurement of individual differences in EF (Carlson, 2005), we used some of those procedures at 26 months. Using a battery of effortful control tasks at 22 and 33 months, Kochanska et al. (2000) found that maternal responsiveness at 22 months, but not at 9 nor 14 months, was related to child effortful control at both 22 and 33 months. Our research shares comparable findings, suggesting that parenting assessed between 9 and 15 months does not relate to impulse control (this study) or effortful control (Kochanska et al., 2000). However, Kochanska et al.'s positive findings with responsiveness assessed at 22 months suggest that this may change as the child gets older or, again, that more proximal measurement points are more strongly related. More longitudinal studies of parenting are needed to shed light on concurrent and prospective links between parenting and child EF at different ages. Furthermore, and in line with recent research, one might speculate that caregiving and child temperament interact in influencing the development of certain aspects of EF (see Crockenberg & Leerkes, 2006; Kochanska, Aksan, & Joy, 2007, for recent examples of such moderating influences on other child outcomes).

A secondary aim of this study was to develop EF measures for younger age groups. We simplified tasks originally designed for 24-month-olds and examined their capacity to detect variation as well as their relations to validated measures at 26 months and to parenting. Despite their low intercorrelation, both 18-month tasks (Hide the Pots and Categorization) were found to be positively related to Conflict at 26 months, and Hide the Pots was associated with later Impulse Control as well. The magnitude of these longitudinal links was similar to that reported by Hughes and Ensor (2007) between 2 and 3 years of age. This tentatively suggests that both tasks may be valid assessments of early manifestations of EF. However, one may also argue that Hide the Pots is more valid than the sorting task, given that it showed more variability, clearer connections to 26-month tasks, and more compelling links with parenting. Overall, the current adaptation of Spin the Pots can reasonably be considered as a useful first step in developing EF measures for 18-month-olds, but more research is needed before one can confidently conclude that either of our 18-month tasks can be used by itself as an indicator of early EF.

This study provides among the first longitudinal evidence that early parent–child relationships may be useful in understanding the origins and growth of individual differences in child EF. Given that the associations found were derived from observations of different individuals (parent behavior vs. child performance), assessed in different contexts and at different times (with delays of up to 14 months), one can reasonably assume that the links are conceptually robust. Given that parenting contributed substantial additional variance in child EF above its well-documented precursors (child cognitive functioning and maternal education), we argue that further exploration of parenting is likely to be a fruitful path for research on childhood EF. However, intriguing questions remain, for instance, with respect to which aspects of parenting contribute to which components of EF, at which ages, and via which neural and social mechanisms. Given the impressive range of child cognitive and socioemotional outcomes related to EF, the search for the social contexts most likely to favor its development is an important task for future research.

References

Ainsworth, M. D. S., Bell, S. M., & Stayton, D. J. (1974). Infant–mother attachment and social development: Socialization as a product of reciprocal responsiveness to signals. In M. P. M. Richards (Ed.), *The introduction of the child into a social world* (pp. 99–135). Cambridge, UK: Cambridge University Press.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: Freeman.

Bayley, N. (1993). *Bayley Scales of Infant Development* (2nd ed.). New York: Psychological Corporation.

Bernier, A., & Meins, E. (2008). A threshold approach to understanding the origins of attachment disorganization. *Developmental Psychology, 44*, 969–982.

Biederman, J., Monuteaux, M. C., Doyle, A. E., Seidman, L. J., Wilens, T. E., Ferrero, F., et al. (2004). Impact of executive function deficits and ADHD on academic outcomes in children. *Journal of Consulting and Clinical Psychology, 72*, 757–766.

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development, 78*, 647–663.

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental Neuropsychology, 19*, 273–293.

Carlson, S. M. (2003). Executive function in context: Development, measurement, theory, and experience. *Monographs of the Society for Research in Child Development, 68*(Serial No. 274).

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28*, 585–615.

Carlson, S. M., Mandell, D. J., & Williams, L. (2004). Executive functions and theory of mind: Stability and prediction from age 2 to 3. *Developmental Psychology, 40*, 1105–1122.

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72*, 1032–1053.

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development, 11*, 73–92.

Carlson, S. M., & Wang, T. S. (2007). Inhibitory control and emotion regulation in preschool children. *Cognitive Development, 22*, 489–510.

Chasiotis, A., Kiessling, F., Hofer, J., & Campos, D. (2006). Theory of mind and inhibitory control in three cultures: Conflict inhibition predicts false belief understanding in Germany, Costa Rica and Cameroon. *International Journal of Behavioral Development, 30*, 249–260.

Chugani, H. T., Behen, M. E., Muzik, O., Juhasz, C., Nagy, F., & Chugani, D. C. (2001). Local brain functional activity following early deprivation: A study of postinstitutionalized Romanian orphans. *Neuroimage, 14*, 1290–1301.

Clark, C., Prior, M., & Kinsella, G. (2002). The relationship between executive function abilities, adaptive behaviour, and academic achievement in children with externalising behaviour problems. *Journal of Child Psychology and Psychiatry, 43*, 785–796.

Crockenberg, S. C., & Leerkes, E. M. (2006). Infant and maternal behavior moderate reactivity to novelty to predict anxious behavior at 2.5 years. *Development and Psychopathology, 18*, 17–34.

Curtis, W. J., & Cicchetti, D. (2007). Emotion and resilience: A multilevel investigation of hemispheric electroencephalogram asymmetry and emotion regulation in maltreated and nonmaltreated children. *Development and Psychopathology, 19*, 811–840.

De Bellis, M. D. (2001). Developmental traumatology: The psychobiological development of maltreated children and implication for research, treatment and policy. *Development and Psychopathology, 13*, 539–564.

De Bellis, M. D. (2005). The psychobiology of neglect. *Child Maltreatment, 10*, 150–172.

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science, 318*, 1387–1388.

Diamond, A., & Goldman-Rakic, P. S. (1989). Comparison of human infants and rhesus monkeys on Piaget's AB task: Evidence for dependence on dorsolateral prefrontal cortex. *Experimental Brain Research, 74*, 24–40.

Espy, K., McDiarmid, M. M., Cwik, M. F., Stalets, M. M., Hamby, A., & Senn, T. E. (2004). The contribution of executive functions to emergent mathematic skills in preschool children. *Developmental Neuropsychology, 26*, 465–486.

Francis, D. D., & Meaney, M. J. (1999). Maternal care and the development of stress responses. *Current Opinion in Neurobiology, 9*, 128–134.

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134*, 31–60.

Glaser, D. (2000). Child abuse and neglect and the brain: A review. *Journal of Child Psychology and Psychiatry, 41*, 97–116.

Golden, C. J. (1981). The Luria-Nebraska Children's Battery: Theory and formulation. In G. W. Hynd & J. E. Obrzut (Eds.), *Neuropsychological assessment and the school-aged child: Issues and procedures* (pp. 277–302). New York: Grune & Stratton.

Greenough, W. T., & Black, J. E. (1992). Induction of brain structure by experience: Substrates for cognitive development. In M. R. Gunnar & C. A. Nelson (Eds.), *Developmental behavioral neuroscience. The Minnesota* *symposia on child psychology* (Vol. 24, pp. 155–200). Hillsdale, NJ: Erlbaum.

Grolnick, W., Frodi, A., & Bridges, L. (1984). Maternal control style and the mastery motivation of one-year-olds. *Infant Mental Health Journal, 5*, 72–82.

Grolnick, W. S., Gurland, S. T., DeCourcey, W., & Jacob, K. (2002). Antecedents and consequences of mother's autonomy support: An experimental investigation. *Developmental Psychology, 38*, 143–155.

Grolnick, W. S., & Ryan, R. M. (1989). Parent styles associated with children's self-regulation and competence in school. *Journal of Educational Psychology, 81*, 143–154.

Grossmann, K. E., & Grossmann, K. (1991). Attachment quality as an organizer of emotional and behavioral responses in a longitudinal perspective. In C. M. Parkes, J. Stevenson-Hinde, & P. Marris (Eds.), *Attachment across the life cycle* (pp. 93–144). New York: Tavistock/Routledge.

Gunnar, M. R., & Donzella, B. (2002). Social regulation of the cortisol levels in early human development. *Psychoneuroendocrinology, 27*, 199–220.

Gunnar, M. R., Fisher, P. A., & the Early Experience, Stress and Prevention Network. (2006). Bringing basic research on early experience and stress neurobiology to bear on preventive interventions for neglected and maltreated children. *Development and Psychopathology, 18*, 651–677.

Hane, A. A., & Fox, N. A. (2006). Ordinary variations in maternal caregiving influence human infants' stress reactivity. *Psychological Science, 17*, 550–556.

Harrist, A. W., & Waugh, R. M. (2002). Dyadic synchrony: Its structure and function in children's development. *Developmental Review, 22*, 555–592.

Hofer, M. A. (1995). Hidden regulators: Implications for a new understanding of attachment, separation, and loss. In S. Goldberg, R. Muir & J. Kerr (Eds.), *Attachment theory: Social, development and clinical perspectives* (pp. 203–230). Hillsdale, NJ: Analytic Press.

Hofer, M. A. (1996). Multiple regulators of ultrasonic vocalization in the infant rat. *Psychoneuroendocrinology, 21*, 203–217.

Hughes, C. (2002). Executive functions and development: Why the interest? *Infant and Child Development, 11*, 69–71.

Hughes, C., Dunn, J., & White, A. (1998). Trick or treat? Uneven understanding of mind and emotion and executive dysfunction in ''hard-to-manage'' preschoolers. *Journal of Child Psychology and Psychiatry, 39*, 981–994.

Hughes, C., & Ensor, R. (2005). Executive function and theory of mind in 2 year-olds: A family affair? *Developmental Neuropsychology, 28*, 645–668.

Hughes, C., & Ensor, R. (2007). Executive function and theory of mind: Predictive relations from ages 2 to 4. *Developmental Psychology, 43*, 1447–1459.

Huttenlocher, P. R. (2002). *Neural plasticity: The effects of environment on the development of the cerebral cortex*. Cambridge, MA: Harvard University Press.

Kinsella, G., Prior, M., Sawyer, M., Ong, B., Murtagh, D., Eisenmajer, R., et al. (1997). Predictors and indicators of academic outcome in children 2 years following traumatic brain injury. *Journal of the International Neuropsychological Society, 3*, 608–616.

Kochanska, G., Aksan, N., & Joy, M. E. (2007). Children's fearfulness as a moderator of parenting in early socialization: Two longitudinal studies. *Developmental Psychology, 43*, 222–237.

Kochanska, G., Murray, K. T., & Harlan, E. T. (2000). Effortful control in early childhood: Continuity and change, antecedents, and implications for social development. *Developmental Psychology, 36*, 220–232.

Kochanska, G., Murray, K. T., Jacques, T. Y., Koenig, A. L., & Vandegeest, K. A. (1996). Inhibitory control in young children and its role in emerging internalization. *Child Development, 67*, 490–507.

Kopp, C. (1982). Antecedents of self-regulation: A developmental perspective. *Developmental Psychology, 18*, 199–214.

Kraemer, G. (1992). A psychobiological theory of attachment. *Behavioral and Brain Sciences, 15*, 493–541.

Landry, S. H., Garner, P. W., Swank, P. R., & Baldwin, C. D. (1996). Effects of maternal scaffolding during joint toy play with preterm and full-term infants. *Merrill-Palmer Quarterly, 42*, 177–199.

Landry, S. H., Smith, K. E., Swank, P. R., & Miller-Loncar, C. L. (2000). Early maternal and child influences on children's later independent cognitive and social functioning. *Child Development, 71*, 358–375.

Lemelin, J.-P., Tarabulsy, G. M., & Provost, M. A. (2006). Predicting preschool cognitive development from infant temperament, maternal sensitivity, and psychosocial risk. *Merrill-Palmer Quarterly, 52*, 779–806.

Marshall, P. J., & Fox, N. A. (2004). A comparison of the electroencephalogram between institutionalized and community children in Romania. *Journal of Cognitive Neuroscience, 16*, 1327–1338.

McEwen, B. S. (1999). The effects of stress on structural and functional plasticity in the hippocampus. In D. S. Charney, E. J. Nestler & B. S. Bunney (Eds.), *Neurobiology of mental illness* (pp. 475–493). New York: Oxford University Press.

Meins, E., Fernyhough, C., Fradley, E., & Tuckey, M. (2001). Rethinking maternal sensitivity: Mother's comments on infant's mental processes predict security of attachment at 12 months. *Journal of Child Psychology and Psychiatry, 42*, 637–648.

Meins, E., Fernyhough, C., Wainwright, R., Gupta, M. D., Fradley, E., & Tuckey, M. (2002). Maternal mind-mindedness and attachment security as predictors of theory of mind understanding. *Child Development, 73*, 1715–1726.

Mezzacappa, E. (2004). Alerting, orienting, and executive attention: Developmental properties and sociodemographic correlates in an epidemiological sample of young, urban children. *Child Development, 75*, 1373–1386.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contribution to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41,* 49–100.

Moran, G., Forbes, L., Evans, E., Tarabulsy, G. M., & Madigan, S. (2008). Both maternal sensitivity and atypical maternal behavior independently predict attachment security and disorganization in adolescent mother-infant relationships. *Infant Behavior and Development, 31,* 321–325.

Müller, U., Zelazo, P. D., & lmrisek, S. (2005). Executive function and children's understanding of false belief: How specific is the relation? *Cognitive Development, 20,* 173–189.

Nelson, C. A. (2000). The effects of early adversity on neurobehavioral development. *The Minnesota symposia on child psychology* (Vol. 31). Mahwah, NJ: Erlbaum.

Nelson, C. A., & Bloom, F. E. (1997). Child development and neuroscience. *Child Development, 68,* 970–987.

Nelson, C. A., Thomas, K. M., & de Haan, M. (2006). Neural bases of cognitive development. ln D. Kuhn, R. S. Siegler, W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology: Vol. 2. cognition, perception, and language* (6th ed., pp. 3–57). Hoboken, NJ: Wiley.

Pederson, D. R., Gleason, K. E., Moran, G., & Bento, S. (1998). Maternal attachment representations, maternal sensitivity, and the infant-mother attachment relationship. *Developmental Psychology, 34,* 925–933.

Pederson, D. R., & Moran, G. (1995). A categorical description of infant-mother relationships in the home and its relation to Q-sort measures of infant-mother interaction. ln E. Waters, et al. (Eds.), New growing points of attachment theory and research. *Monographs of the Society for Research in Child Development, 60*(Serial No. 244).

Pederson, D. R., Moran, G., Sitko, C., Campbell, K., Ghesquire, K., & Acton, H. (1990). Maternal sensitivity and the security of infant-mother attachment: A Q-sort study. *Child Development, 61,* 1974–1983.

Perner, J., Lang, B., & Kloo, D. (2002). Theory of mind and self-control: More than a common problem of inhibition. *Child Development, 73,* 752–767.

Rothbart, M. K. (1989). Temperament and development. ln G. Kohnstamm, J. Bates & M. K. Rothbart (Eds.), *Temperament in childhood* (pp. 187–248). Chichester, UK: Wiley.

Rueda, M. R., Posner, M. I., & Rothbart, M. K. (2005). The development of executive attention: Contributions to the emergence of self-regulation. *Developmental Neuropsychology, 28,* 573–594.

Rutter, M., O'Connor, T., & the English and Romanian Adoptee Study Team. (2004). Are there biological programming effects for psychological development? Findings from a study of Romanian adoptees. *Developmental Psychology, 40,* 81–94.

Rutter, M., & the ERA Research Team. (1998). Developmental catch-up and deficit following adoption after severe global early privation. *Journal of Child Psychology and Psychiatry, 39,* 465–476.

Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The development of executive functioning and theory of mind: A comparison of Chinese and U.S. preschoolers. *Psychological Science, 17,* 74–81.

Schore, A. N. (1996). The experience-dependant maturation of a regulatory system in the orbital prefrontal cortex and the origin of developmental psychopathology. *Development and Psychopathology, 8,* 59–87.

Schore, A. N. (2001). Effects of a secure attachment relationship on right brain development, affect regulation, and infant mental health. *Infant Mental Health Journal, 22,* 7–66.

Simonds, J., Kieras, J. E., Rueda, M. R., & Rothbart, M. K. (2007). Effortful control, executive attention, and emotional regulation in 7-10-year-old children. *Cognitive Development, 22,* 474–488.

Singer, W. (1995). Development and plasticity of cortical processing architectures. *Science, 270,* 758–764.

Spangler, G., Schieche, M., Ilg, U., Maier, U., & Ackermann, C. (1994). Maternal sensitivity as an external organizer for biobehavioral regulation in infancy. *Developmental Psychobiology, 27,* 425–437.

Stuss, D. T., & Benson, D. F. (1986). *The frontal lobes.* New York: Raven Press.

Tardif, T., Wing-Chee So, C., & Kaciroti, N. (2007). Language and false belief: Evidence for general, not specific, effects in Cantonese-speaking preschoolers. *Developmental Psychology, 43,* 318–340.

Uzgiris, I. C., & Hunt, J. Mc. V. (1975). *Assessment in infancy: Ordinal scales of psychological development.* Champaign: University of Illinois Press.

van der Sluis, S., de Jong, P. F., & van der Leij, A. (2008). Executive functioning in children, and its relations with reasoning, reading, and arithmetic. *Intelligence, 35,* 427–449.

Zelazo, P. D., Carlson, S. M., & Kesek, A. (2008). The development of executive function in childhood. ln C. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (2nd ed., pp. 553–574). Cambridge, MA: MIT Press.

Zelazo, P. D., Carter, A., Reznick, J. S., & Frye, D. (1997). Early development of executive function: A problem-solving framework. *Review of General Psychology, 1,* 198–226.

Zelazo, P. D., Müller, U., Frye, D., & Marcovitch, S. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development, 68*(Serial No. 274).

Child Development, January/February 2010, Volume 81, Number 1, Pages 340–356

Infant Pathways to Externalizing Behavior: Evidence of Genotype × Environment Interaction

Leslie D. Leve
Oregon Social Learning Center

David C. R. Kerr
Oregon Social Learning Center and
Oregon State University

Daniel Shaw
University of Pittsburgh

Xiaojia Ge
University of Minnesota, Twin Cities

Jenae M. Neiderhiser
The Pennsylvania State University

Laura V. Scaramella
University of New Orleans

John B. Reid
Oregon Social Learning Center

Rand Conger
University of California, Davis

David Reiss
Yale University

To further the understanding of the effects of early experiences, 9-month-old infants were observed during a frustration task. The analytical sample was composed of 348 linked triads of participants (adoptive parents, adopted child, and birth parent[s]) from a prospective adoption study. It was hypothesized that genetic risk for externalizing problems and affect dysregulation in the adoptive parents would independently and interactively predict a known precursor to externalizing problems: heightened infant attention to frustrating events. Results supported the moderation hypotheses involving adoptive mother affect dysregulation: Infants at genetic risk showed heightened attention to frustrating events only when the adoptive mother had higher levels of anxious and depressive symptoms. The Genotype × Environment interaction pattern held when substance use during pregnancy was considered.

The independent contributions of genetic and environmental influences to child behavior have been well established over the past several decades using behavioral genetic designs (Reiss, Neiderhiser, Hetherington, & Plomin, 2000; Rutter, 2006). This body of research has indicated moderate genetic effects on most dimensions of externalizing behavior (Rhee & Waldman, 2002). Emanating from and building upon this work, there is a burgeoning body of evidence from behavioral and molecular genetic studies suggesting that environmental factors moderate or mediate genetic influences on

---

This article may not exactly replicate the final version published in the APA journal. It is not the copy of record.

This project was supported by Grant R01 HD042608, NICHD, NIDA, and the Office of the Director, NIH, U.S. PHS (PI Years 1–5: David Reiss, MD; PI Years 6–10: Leslie Leve, PhD). The writing of this manuscript was partially supported by P20 DA017592 and P30 DA023920, NIDA, NIH, U.S. PHS. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Eunice Kennedy Shriver National Institute of Child Health and Human Development or the National Institutes of Health. We would like to thank the birth and adoptive parents who participated in this study and the adoption agency staff members who helped with the recruitment of study participants. Thanks also are due to Matthew Rabel and Adrienne van der Valk for editorial assistance. Special gratitude is given to Remi Cadoret and Beverly Fagot who were centrally involved in this work prior to their deaths.

Correspondence concerning this article should be addressed to Leslie D. Leve, Oregon Social Learning Center, 10 Shelton McMurphey Boulevard, Eugene, OR 97401. Electronic mail may be sent to lesliel@oslc.org.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8101-0022

problem behaviors. Studies of Genotype × Environment (G × E) interaction help to illustrate how inherited characteristics render some individuals more susceptible to the negative or positive effects of specific environments (Moffitt, Caspi, & Rutter, 2005). This line of investigation is important in helping to identify mechanisms whereby specific environmental processes might offset or exacerbate genetic risk, thereby suggesting potential targets for preventive interventions. Most of the published studies in this area, however, have used retrospective measures of the environment or have studied environmental processes in middle childhood and beyond (Caspi et al., 2002; Feinberg, Button, Neiderhiser, Reiss, & Hetherington, 2007). Such approaches miss the opportunity to capture environmental experiences during the early childhood developmental period, which has been found to be malleable to preventive interventions (Olds, 2002; Shaw, Dishion, Supplee, Gardner, & Arnds, 2006). In this study, we examine specific genetic and family environmental processes on a known precursor of externalizing behavior: infant heightened attention to frustrating events (Crockenberg, Leerkes, & Barrig Jo, 2008).

Connection Between Attention to Frustration and Subsequent Externalizing Behavior

The focus on infant attention to frustrating events, which has known links to subsequent externalizing behavior, was selected because infants are not developmentally capable of engaging in behaviors typically captured by standard definitions of externalizing behavior (e.g., cheating, lying, or directed aggression). Further, temperamental characteristics such as anger do not show a consistent relation with externalizing problems until 12–24 months of age (Blair, 2002; Lyons-Ruth, Easterbrooks, & Cibelli, 1997; Smeekens, Riksen-Walraven, & van Bakel, 2007). The behaviors involved in attention regulation undergo a developmental shift between 3 and 6 months of age (e.g., the acquisition of visual orientation skills and the ability to disengage visual gaze; Johnson, Posner, & Rothbart, 1991; Posner & Petersen, 1990), making infancy an ideal time to examine attentional processes. One aspect of attention—infant's attention to frustrating events (defined in this article as focus and persistence on a frustrating event)—has been examined in numerous studies during early childhood. Stifter and Braungart (1995) indicated that 5- and 10-month-olds were more likely to visually shift their attention away from a frustrating event when infant distress was decreasing, suggesting a regulatory role of infant ability to shift attention away from frustrating events. In addition, converging evidence from longitudinal studies has indicated a link between attentional focus on frustrating events and later externalizing problems.

In a longitudinal study of 5-month-old infants, Crockenberg et al. (2008) found that infants who failed to shift attention away from frustrating events (an arm restraint and a toy-removal task) exhibited increased aggressive behavior at 2½ years of age. Attention to frustration was defined by coder ratings of infant attempts to inspect the frustrating event and situation. The association between this infant behavior and later aggression was the strongest when mothers encouraged their child to look at the frustrating event, which is consistent with the notion that maternal regulation processes moderate pathways to externalizing problems. Crockenberg et al. suggested that attention to a frustrating event is the dominant response and that successfully looking away reflects the ability to adaptively manage frustration. Regulation characteristics of the mother might undermine the development of a child's capacity to shift attention away from frustrating events, inadvertently contributing to the development of externalizing problems. In the present study, we mirrored Crockenberg et al.'s work by utilizing a similar toy-removal task across repeated intervals and by coding for infant attention to the frustrating event and situation.

Although we are unaware of other studies that have employed toy-removal protocols to measure attention to frustrating events, several longitudinal studies have examined early childhood attentional problems using other methods and have found similar associations with subsequent externalizing problems. For example, Caspi, Henry, McGee, Moffitt, and Silva (1995) found that observer ratings of 3-year-olds' inability to modulate attention and lack of control were prospectively associated with externalizing problems in adolescence. Further, at 21 years of age, young children with these attention problems were more likely to be diagnosed with antisocial personality disorder (Caspi, Moffitt, Newman, & Silva, 1996). Tracing the pathways to externalizing problems earlier in development in two independent cohorts of low-income families, Shaw, Keenan, and Vondra (1994) and Shaw et al. (1998) observed mothers with their 12-month-olds during a task designed to elicit infant frustration behavior (i.e., high chair task). In both cohorts, boys who were higher on persistence in the face of caregiver unresponsivity at 12 months showed elevated rates of externalizing problems at age 2 and between ages 3 and 3½ based on observations and maternal ratings of disruptive behavior (e.g., noncompliance, aggression). These studies replicated earlier work by Martin (1981), who used the same high chair task and found remarkably similar results for boys from less disadvantaged backgrounds.

There is also evidence to indicate that at-risk children may have heightened states of arousal and reactivity (Fisher, Stoolmiller, Gunnar, & Burraston, 2007), which might cause them to focus on negative cues that typically invoke negative emotions such as anger (Shackman, Shackman, & Pollak, 2007). Consistent with this body of work, we hypothesize that, under conditions of frustration, infants at risk for externalizing problems will react with increased attention and persistence. The extent to which the early home environment and genetic factors jointly affect the development of attention to frustration has yet to be investigated; in this study, we extended prior work by focusing on genetic and environmental influences on infants' attention behavior during a frustration task.

Genetic Contributions to Externalizing Behavior

The results from twin and adoption studies on the heritability of externalizing behavior provide evidence that the externalizing behaviors of interest in the proposed study are genetically influenced. In a meta-analysis of 24 genetically informative studies on aggression, Miles and Carey (1997) found that genetic effects accounted for approximately 50% of the variance in aggressive behavior. More recently, in a meta-analysis of 51 twin and adoption studies of antisocial behavior, Rhee and Waldman (2002) found more modest genetic influences (41%). The findings from recent twin and adoption studies also point to genetic influences on attention deficit problems (Thapar, O'Donovan, & Owen, 2005), novelty seeking (Hiroi & Agatsuma, 2005), and the associations between novelty seeking, externalizing behavior, and substance use (Laucht et al., 2007). Although useful in identifying potentially heritable aspects of externalizing behavior, such findings provide little information about the mechanisms of development, how the phenotype might evolve across development, and how genetic aspects of externalizing behavior during adolescence and adulthood relate to the genetic aspects of infants' propensity for externalizing behavior.

The findings from life-course studies by Laucht and colleagues have provided information about possible mechanisms in this regard. In this work, novelty seeking was found to be associated with a polymorphism of the DRD4 gene (Laucht, Becker, El-Faddagh, Hohm, & Schmidt, 2005). In a longitudinal study from infancy to adolescence, Laucht, Becker, and Schmidt (2006) found that visual exploratory behavior (i.e., longer initial looking times) and a higher response decrement (i.e., habituation) in 3-month-olds was associated with novelty seeking in adolescence and the 7r DRD4 allele variant. This is one of the few areas where a genetically influenced characteristic has been shown to have different and specific phenotypic expressions across development, manifested as attention during early infancy and as novelty seeking (a form of externalizing behavior) during adolescence. Guided by this genetically informed work and the phenotypic work on the association between attention to frustration and later externalizing problems described earlier, we focused on infant attention to frustration to better understand the early mechanisms underlying the development of externalizing behavior.

Role of Parental Affective Dysregulation

According to Crockenberg et al. (2008), maternal behavior plays a key role in the development of a child's attention behavior and externalizing trajectories. One of the most well-documented risk factors during infancy of current and later externalizing problems is maternal depression (Campbell, Matestic, von Stauffenberg, Mohan, & Kirchner, 2007; Downey & Coyne, 1990). There is an extensive body of evidence to suggest that caregiver depression interferes with the ability to sensitively parent (Cicchetti, Rogosch, & Toth, 1998). Similarly, maternal anxiety has been shown to relate to problem behaviors in children, including anxiety and depression, attention deficit disorders, and oppositional defiant disorders (Meadows, McLanahan, & Brooks-Gunn, 2007). Thus, maternal affective dysregulation during infancy is commonly viewed as one of the strongest environmental predictors of child problems. The adoption design provides a unique opportunity to examine whether associations between caregiver affective state and child externalizing problems are found when family members are genetically unrelated.

Current Study

In this study, a prospective adoption design is employed to examine hypotheses regarding the environmental moderation of genetic risk for externalizing behavior. In this design, the infant adopted by nonrelative parents shares 50% of his or her segregating genes with each of the birth parents but none with the adoptive parents. Conversely, the adopted child shares 100% of the rearing environment with the adoptive parents but none of the rearing environment with the birth parents. As such, genetic influences can be inferred based on similarities between the child and the birth parent(s), and environmental influences can be inferred based on similarities between the child and the adoptive parents. This design strength results from the fact that children placed in nonrelative adoptive homes do not share genes with their caregivers; thus, genetic influences cannot account for associations often seen between the caregiving environment and child behavior.

To date, only one other prospective longitudinal adoption study has examined genetic and environmental effects on social processes during infancy: the Colorado Adoption Project (CAP; Plomin, DeFries, & Fulker, 1988). Although the CAP has yielded evidence of G × E interactions on social behaviors during later childhood (Hershberger, 1994; O'Connor, Caspi, DeFries, & Plomin, 2003), G × E interaction effects during infancy have not been published to date. In the present study, we built upon the methods of the CAP by collecting observational assessments of social processes in the adoptive family and by using a fine-tuned assessment protocol with birth parents that included multiple measures of each targeted social construct to increase the reliability of genetic indices. By using a theory- and research-guided approach to carefully match birth parents' social characteristics with hypothesized externalizing and internalizing child outcomes a priori (Leve et al., 2007), we sought to increase the likelihood that G × E interactions would be detected in those domains.

Guided by the research reviewed earlier, we formulated three hypotheses: (a) that genetic effects on pathways to externalizing behavior could be traced intergenerationally via the association between birth parent externalizing behavior and heightened infant attention to frustration; (b) that adoptive parent affective state (e.g., anxious or depressive symptoms) would relate to infant attention to frustration, with higher levels of anxiety and depression being associated with greater infant attention to frustration; and (c) that the affective state of the adoptive parents would moderate genetic influences on infant attention during a frustration task, increasing the association between birth parent externalizing problems and infant attention to frustration. Because substance use during pregnancy can confound genetic effects on development, we examined birth mother prenatal substance use in a secondary set of analyses.

Method

Participants

The sample consisted of participants in the Early Growth and Development Study (EGDS), an ongoing, longitudinal, multisite study of 361 linked triads of adopted children, adoptive parents, and birth parents. The primary goal of the EGDS is to examine the effects of G × E interaction and correlation on the social and emotional development of young children. The recruitment of participants occurred between 2003 and 2006, beginning with the recruitment of adoption agencies ($N = 33$ agencies in 10 states located in the Northwest, Mid-Atlantic, and Southwest regions of the United States). The participating agencies reflected the full range of adoption agencies in the United States: public, private, religious, secular, those favoring open adoptions, and those favoring closed adoptions. Each adoption agency appointed a liaison from their organization to assist with recruitment. The agency liaisons identified participants who completed an adoption plan through their agency and met the study's eligibility criteria: (a) the adoption placement was domestic, (b) the baby was placed within 3 months postpartum, (c) the baby was placed with a nonrelative adoptive family, (d) the baby had no known major medical conditions such as extreme prematurity or extensive medical surgeries, and (e) the birth and adoptive parents were able to read or understand English at the eighth-grade level. The study participants were representative of the adoptive and birth parent populations that completed adoption plans at the participating agencies during the same time period (Leve et al., 2007).

The analytical sample for this report consisted of 348 linked triads from the first wave of assessment from whom at least two of the three members of the triad had data on the variables used in this study. In addition, data from 105 birth fathers (BFs) are included. The children were 9 months old during the assessment, $M = 9.20$ ($SD = 0.55$ months). Four triads were excluded from subsequent analyses on the basis of infant age (i.e., 1 child was 6 months old, and 3 children were more than 12 months old). Forty-three percent of the children were female. The mean child age at the adoption placement was 3 days (*SD* = 5 days). The adoptive families were typically college-educated, middle-class families, and the mean ages of adoptive mothers (AMs) and fathers (AFs) were 37 and 38, respectively. Ninety-three percent of the AMs and 92% of the AFs were Caucasian; 4% of the AMs and 5% of the AFs were African American; 2% of the AMs and AFs were multiethnic; 1% of the AMs and AFs were Hispanic or Latino; and the remaining participants were not identified or were of other ethnic status. The adoptive parents had been married an average of 11.8 years (*SD* = 5.1 years), and 51% of the adoptive families had at least one additional child in the home at the time of the writing of this report. Information about which adoptive parent was the primary caregiver was not collected. The birth parents typically had high school or trade school education levels and household incomes under $25,000, and the mean ages of birth mothers (BMs) and BFs was 24 and 25, respectively. Seventy-eight percent of the BMs and 63% of the BFs were Caucasian; 11% of the BMs and 20% of the BFs were African American; 5% percent of the BMs and BFs were multiethnic; 4% of the BMs and 8% of the BFs were Hispanic or Latino; and the remaining participants were not identified or were of other ethnic status. There were no significant differences in demographic characteristics between the full EGDS sample and the participants in the analytical sample used in this report. Complete demographic information on the full EGDS sample regarding AM, AF, BM, and BF age, race, education, and income are presented in Leve, Neiderhiser, Scaramella, and Reiss (2008).

The study design included a birth parent in-person assessment between 3 and 6 months postpartum and an adoptive family in-person assessment when the child was 9 months old. The birth parent assessment was 2 hr long and was conducted in the home or in another convenient location; the adoptive family assessment was 2½ hr long and was conducted in the home. The participants were paid for their time. For the birth and adoptive parent assessments, the interviewers asked computer-assisted interview questions, and each participant independently completed a set of questionnaires. In addition, the adoptive parents were observed individually (and once conjointly) in a series of interaction tasks with their child. The birth and adoptive parent assessments were completed by separate teams of interviewers to prevent data sharing across members of the adoption triad. The interviewers completed a minimum of 40 hr of training, including a 2-day group session, pilot interviews, and videotaped feedback prior to administering interviews with study participants. All interviews were audio- or video-recorded, and feedback was provided by a trained evaluator for a random selection of 15% of the interviews to ensure adherence to the study's standardized interview protocols. Additional details on the EGDS study recruitment procedures, sample, and assessment methods can be found in Leve et al. (2007).

Measures

Genetic Risk for Externalizing Behavior

To estimate genetic risk for externalizing behavior, we administered three birth parent self-report measures considered to be indicators of externalizing behavior: lifetime alcohol, tobacco, and other drug (ATOD) dependence; delinquency; and novelty seeking.

ATOD dependence. ATOD dependence was measured using a modified version of the Composite International Diagnostic Interview–Short Form (CIDI–SF) Alcohol and Drug Dependence scales (Kessler, Andrews, Mroczek, Ustun, & Wittchen, 1998). The original CIDI–SF questions were modified for this study to pertain to lifetime use (rather than 12-month use) and to include a set of tobacco-dependence questions. The indicators of lifetime problem use of alcohol, tobacco, marijuana, and other drugs were then combined to form a composite ATOD-dependence score. Alcohol dependence was calculated as the sum of seven items from the CIDI–SF assessing alcohol-dependence symptoms in the past 12 months (BM: $\alpha = .82$; BF: $\alpha = .85$). For tobacco dependence, a 4-point scale was created using questions regarding past or current regular tobacco use and two dependence symptoms: 0 (*never used*), 1 (*past or current regular use but other dependence symptoms denied*), 2 (*one dependence symptom endorsed*), and 3 (*both dependence symptoms endorsed*). Marijuana dependence was calculated as the sum of seven CIDI–SF dependence items (BM: $\alpha = .85$, BF: $\alpha = .88$). Other drug dependence included the following eight substance classes: sedatives, tranquilizers, amphetamines, painkillers, inhalants, cocaine, heroin, and hallucinogens. More than half of the birth parents reported use of one or more of these substance classes. However, use within any specific substance class was relatively low (14%–26%). Therefore, we created dependence symptom scales for each substance class (based on the sum of CIDI–SF dependence items for that substance class; BM: α = .63–.97; BF: α = .74–.96). The other drug-dependence indicator was calculated by averaging the standardized dependence symptom scales across each substance class (BM: α = .85; BF: α = .83).

The ATOD dependence indicators were each standardized, and a composite ATOD-dependence score was created (BM: α = .72; BF: α = .71).

Delinquency. Delinquency was measured using the 38-item Elliot Social Behavior Questionnaire (Elliott, Huizinga, & Ageton, 1985). On this self-report delinquency scale, birth parents rated items reflecting their engagement in various delinquent behaviors (e.g., purposely damaged or destroyed property, purposely set a fire, carried a hidden weapon, or stole something). Items were summed to create a delinquency score (BM: α = .88; BF: α = .91). The scales were log-transformed to reduce skewness.

Novelty seeking. Novelty seeking was based on the 20-item Novelty Seeking subscale of the Temperament Character Inventory (BM: α = .73; BF: α = .68; Cloninger, Svaric, & Przybeck, 1993). This measure of general temperament and character is composed of 125 true or false items. Higher scores on the Novelty Seeking subscale indicate an exploratory and novelty seeking personality, impulsive decision making, quick loss of temper, and excitable, enthusiastic, easily bored, and ardent behavior.

Associations among the externalizing indicators. For the BMs, the three standardized indicators (ATOD dependence, delinquency, and novelty seeking) were significantly correlated (r = .28–.41, p < .001) and each loaded significantly on a one-factor solution (.76, .79, and .69, respectively), forming an internally consistent scale (α = .61). The BF indicators did not show the same level of internal consistency, with a nonsignificant correlation among one set of variables and a marginally internally consistent scale (α = .56). Still, the three indicators all loaded significantly on a one-factor solution (.73, .85, and .59, respectively). The three BM externalizing behavior indicators were used in structural equation modeling (SEM), and a factor score was created for use in graphing. A BF externalizing behavior factor score based on the 99 BFs with data on all three indicators was calculated and used in the BF analyses.

Adoptive Parent Affective State

Adoptive parent affective state was examined using the Beck Anxiety Inventory (BAI; Beck & Steer, 1993a) and the Beck Depression Inventory (BDI; Beck & Steer, 1993b). Anxiety symptoms were calculated as the sum on the 21-item BAI (AM: α = .76; AF: α = .74). Depressive symptoms were calculated as the sum of 20 items from the 21-item BDI (AM: α = .72; AF: α = .75); the suicidal ideation item was not administered in the present study to minimize situations where clinical follow-up would be necessitated. The anxiety and depressive scales were log-transformed to reduce skewness. The resulting scales were correlated for AMs (r = .48) and AFs (r = .54) but were not correlated between parents. Thus, we standardized the anxiety and depression scales for each parent based on the means. Higher scores indicated greater anxious and depressive symptoms.

Infant Attention During a Frustration Task

Infant attention to frustration was measured during a barrier task as part of the laboratory temperament assessment battery assessment protocols (Goldsmith & Rothbart, 1996). In this video-recorded task, the child is seated in a high chair while six trials are run in 30-s intervals, alternating between neutral trials (Trials 1, 3, and 5) and frustration trials (Trials 2, 4, and 6). During each neutral trial, the same stimulating toy is presented on a high chair tray to the infant for 30 s. On each frustration trial, a clear Plexiglas barrier is placed between the infant and the toy for 30 s so that the infant can readily see the toy but cannot touch it. The barricade task was selected because of its developmental appropriateness and its similarity to attention to frustration tasks used in other studies during infancy (Crockenberg et al., 2008).

Infant attention during each trial of the barricade task was coded from video-recorded data using the interest and persistence scale from the family interaction behavior codes (Dogan et al., 2005). This scale assesses a child's level of engagement with the toy provided in the task. The primary measure is the amount of time the child engages with or attempts to engage with the toy. It also includes the amount of interest the child initially displays in the toy during each interval. A child scoring high on interest shows attention to, focus on, and thorough attempts to explore the toy provided to him or her for an extended period of time. Trained coders rated infants on interest behavior during each of the six trials using a 9-point Likert-type scale: 1 (*not at all characteristic*) to 9 (*mainly characteristic*). Thirty-five percent of the infants were dual-coded by independent coders to establish reliability. Coder intraclass correlations ranged from .91 to .97 across the six trials, indicating high inter rater reliability.

A factor analysis with oblique rotation supported a two-factor solution and a distinction between the trial types; neutral trials loaded .76–.84 on one factor, while frustration trials loaded .77–.86 on a second factor (r between factors = .315, $p < .001$). Attention ratings were standardized at each trial. The measure of attention during neutral trials was calculated as the mean of standardized ratings on Trials 1, 3, and 5 ($\alpha = .71$), and attention during frustration trials was calculated as the mean on Trials 2, 4, and 6 ($\alpha = .76$). Attention during the frustration trials served as the dependent measure of interest for the current analyses; attention during the neutral trials was included as a control variable to take into account individual differences in general attention levels during neutral situations.

Additional Control Variables

Several additional control variables were included to control for phenomena noncentral to the study hypotheses.

Openness in adoption. To control for similarities between birth and adoptive families resulting from contact and knowledge between parties, we controlled for the level of openness in the adoption. We measured openness in adoption using a composite index of BM, BF, AM, and AF perceived openness using a 7-point scale: 1 (*very closed*) to 7 (*very open*). Interrater agreement was high (r range = .66–.81; Ge et al., 2008).

Child sex. Child sex was coded as 1 (*female*) or 2 (*male*).

BM prenatal ATOD use. Maternal prenatal ATOD use can confound estimates of the genetic influences related to substance use (such as our ATOD dependence measure, described above). Thus, prenatal ATOD use was included as a control variable in a secondary set of analyses. The BMs reported prenatal ATOD use on each of 10 substance classes (tobacco, alcohol, sedatives, tranquilizers, amphetamines, painkillers, inhalants, cocaine, heroin, and hallucinogens) using a pregnancy history calendar method developed for the study. The BMs were asked to identify key events throughout the pregnancy period (e.g., birthdays, holidays, and vacations) and to note them on a one-page life history calendar, which has been shown to be an effective and reliable tool for recording retrospective events and conditions (Caspi, Moffitt, Thornton, et al., 1996). The BMs then responded to questions about their substance use during the pregnancy period, using the calendar as a reference to facilitate recall.

For tobacco use, the BMs reported whether they had smoked cigarettes in the first, second, and/or third trimesters of this pregnancy using a yes or no format; 59.6% reported not smoking, and 26.2% reported smoking during all three trimesters. For the remaining substances classes, we used a 5-point scale for the 9-month pregnancy period: 0 (*never*) to 4 (*regularly or most days, or every day*). Few BMs reported more than rare alcohol use (5.4%) or any level of use within the other drug classes (ranging from 0.3% for hallucinogens to 16.4% for marijuana), but 27.0% of the mothers reported use of at least one of these substances. Given the skewed distributions of the use indicators, we considered a variety of rescaling options when constructing an overall scale. Dichotomizing all 10 indicators—0 (*no use*) or 1 (*any use*)—yielded an internally consistent scale, Cronbach's α (KR-20) = .67. Hallucinogen use was omitted during scale development due to low frequency but was included in the final scale because of face validity. The sum of dichotomous indicators was positively skewed and was therefore collapsed into a 5-point scale: 0 (*prenatal use of no substances*) to 4 (*prenatal use of four or more substances*).

Data Analysis Approach

The analyses were aimed at examining genetic, family environment, and Genotype × Family Environment effects on infant heightened attention during the frustration task. Significant correlations between an infant characteristic and a birth parent characteristic belonging to the same underlying construct were inferred to reflect genetic influences.

SEM using Mplus (Muthén & Muthén, 2006) was used to test whether BM externalizing behavior (a latent factor indexed by multiple indicators), AM and AF affective state (observed variables), and the interaction between them predicted infants' attention during the frustration trials (an observed variable). We tested the interactions involving AM and AF affective state separately, while controlling for the other adoptive parent's state (e.g., BM Externalizing × AM Affective State, controlling for AF affective state). Three control variables were included in the models: infant sex, openness in adoption, and infant attention during the neutral trials. In addition, we examined BM substance use during pregnancy in follow-up analyses to determine whether genetic influences were confounded by prenatal ATOD use.

All of the variables were standardized prior to modeling and used the maximum likelihood algorithm with robust standard errors (MLR; Muthén & Muthén, 2006) to derive parameter estimates in the presence of missing data. Given the lack of established fit statistics for models involving statistical interaction with latent variables, we followed a series of steps to test these models. First, to test whether the model provided an adequate fit without the interaction term, we computed an MLR chi-square (asymptotically equivalent to Yuan–Bentler T2 χ^2; Muthén & Muthén, 2006) and examined root mean square error of approximation (RMSEA) and comparative fit index (CFI). Second, we used the Satorra–Bentler log-likelihood difference chi-square test to compare a model with the path from the interaction term to the outcome to the model without the path, as one is mathematically nested within the other. Third, when model fit was improved and an individual interaction term was significant, we examined the direction of effects by splitting the sample at the mean of the moderating variable and examining the effects of BM externalizing problems on infant attention to frustration within the resulting subsamples. We used the mean-split approach rather than splitting the sample 1 *SD* above and below the mean due to the small sample sizes that resulted in the latter approach. Nonetheless, we ran preliminary follow-up analyses splitting the sample at 1 *SD* above and below the mean and at 0.5 *SD* above and below the mean to examine comparability of the results using the different approaches.

The same set of interactions was also tested using the reduced BF subsample. We could not use a latent BF externalizing problem variable in the SEM due to the smaller sample size and relatively low association between two of the three ATOD use variables. Thus, there was no advantage to testing the hypotheses in this subsample using SEM. Instead, we used linear regression analyses in SPSS. We examined a BF externalizing problem factor score (listwise $n = 99$), AM affective state, and the interaction between these two variables as predictors of infant attention during the frustration trials. A parallel regression model in which BF externalizing problems interacted with AF affective state was considered separately.

Results

We first examined the means, standard deviations, and bivariate correlations among study variables. Next, we performed SEM to test the hypothesis that BM characteristics would interact with adoptive parent affective state to predict infant heightened attention during the frustration task. As a subsidiary analysis, we examined this hypothesis using a regression approach with the BF data. Finally, we explored prenatal substance use effects as a potential mediating pathway.

Descriptive Statistics and Correlations

The analytic sample consisted of 348 linked triads for whom data were available on at least two of the three members of the triad. Means, standard deviations, correlations among study variables, and *n* sizes for each variable are summarized in Table 1. Infant attention during the frustration trials was moderately associated with infant attention during the neutral trials ($r = .33$, $p < .001$). The correlation matrix also indicated significant associations among the three BM externalizing measures, significant associations between two of the three BF externalizing measures, and significant associations among about half of the BM–BF externalizing behavior cross correlations. In addition, there were several modest but significant correlations among the remaining variables. For example, BM prenatal ATOD use was correlated with the BM externalizing measures and with two of the BF externalizing measures. BM delinquency was also modestly associated with attention during the frustration trials ($r = .13$, $p < .05$) and was inversely associated with AF affective state ($r = -.12$, $p < .05$), and AM affective state was associated with openness ($r = .12$, $p < .05$). Otherwise, the study variables were no more associated than would be expected by chance.

Evaluation of BM × Adoptive Parent Effects

The SEM analyses used full information maximum likelihood, which has been shown to provide unbiased estimates when data are missing at random (Arbuckle, 1996) and to offer greater statistical efficiency compared to mean-imputation, listwise, and pairwise deletion methods (Wothke, 2000). The first series of models tested infant attention during the frustration trials regressed on the latent BM externalizing behavior variable (ATOD dependence, delinquency, and novelty seeking), AM affective state, and their interaction; AF affective state, adoption openness, infant attention during neutral trials, and infant sex were statistically controlled.

Table 1

Means, Standard Deviations, and Correlations Among Study Variables

	1	2	3	4	5	6	7	8	9	10	11	12
1. Attention—frustration trials	—											
2. Attention—neutral trials	.33***	—										
3. Affective state (AF)	−.07	.01	—									
4. Affective state (AM)	.09	−.01	.05	—								
5. ATOD use (BF)	.09	−.05	−.09	.08	—							
6. Delinquency (BF)	.04	.08	−.06	−.06	.44***	—						
7. Novelty seeking (BF)	.00	−.10	.13	−.12	.11	.33**	—					
8. ATOD use (BM)	.11	−.07	−.06	.01	.60***	.25*	−.11	—				
9. Delinquency (BM)	.13*	.01	−.12*	−.02	.33**	.27**	.08	.41***	—			
10. Novelty seeking (BM)	.02	.04	−.03	.01	.15	.15	.06	.28***	.31***	—		
11. Prenatal ATOD use (BM)	.04	−.05	−.04	−.01	.52***	.23*	.01	.56***	.38***	.15**	—	
12. Openness in adoption	.03	.03	−.01	.12*	.02	.05	−.07	.10	−.01	.08	.08	—
M	3.04	5.23	0.00^a	0.00^a	0.00^b	47.28	12.24	0.00^b	41.93	10.57	1.01	4.74
SD	1.16	1.48	1.00	1.00	1.00	9.09	3.50	1.00	5.92	3.84	1.17	1.17
N	312	314	336	348	103	105	104	338	340	340	347	348

Note. AF = adoptive father; AM = adoptive mother; BF = birth father; BM = birth mother; ATOD = alcohol, tobacco, and other drugs. ^{a}Mean of standardized Beck Anxiety Inventory (BAI) and modified Beck Depression Inventory (BDI) scores; unstandardized *M* (*SD*) BAI scores for adoptive fathers and mothers were 3.03 (3.09) and 3.84 (3.55), respectively; adoptive father and mother unstandardized *M* (*SD*) BDI scores were 2.89 (3.37) and 3.55 (3.11), respectively. ^{b}Composite score based on four standardized indicators.
$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

The first of these models (prior to entry of the interaction term) showed an acceptable fit to the data. Higher levels of AM anxious and depressive symptoms were modestly associated with higher levels of infant attention during the frustration trials ($p < .05$). Higher levels of BM externalizing behavior indicated a trend toward being significantly associated with higher levels of infant attention during the frustration trials ($p < .10$). Other than infant attention during the neutral trials ($p < .001$), none of the control variables was associated with infant attention to frustration in this model. Next, a model in which the interaction between BM externalizing problems and AM affective state was fitted to the data. The results are summarized in Figure 1. Compared to the first model, in which this path was fixed at zero, this model fit the data significantly better, Satorra–Bentler χ^2 difference ($df = 1$) = 6.67, $p < .05$. More importantly, the interaction term also was significant ($p < .05$). The AM affective state remained significant ($\beta = .10$, $p < .05$), and the BM externalizing behavior path became significant in the interaction model ($\beta = .17$, $p < .05$).

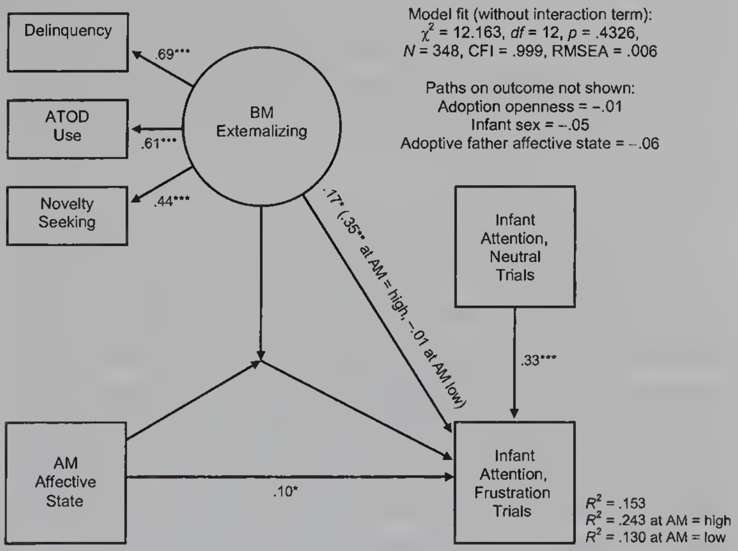

Delinquency
ATOD Use
Novelty Seeking
BM Extemalizing
.69***
.61***
.44***
Model fit (without interaction term):
χ^2 = 12.163, df = 12, p = .4326,
N = 348, CFI = .999, RMSEA = .006
Paths on outcome not shown:
Adoption openness = −.01
Infant sex = −.05
Adoptive father affective state = −.06
.17* (.35** at AM = high, −.01 at AM low)
Infant Attention, Neutral Trials
.33***
AM Affective State
.10*
Infant Attention, Frustration Trials
R^2 = .153
R^2 = .243 at AM = high
R^2 = .130 at AM = low

Figure 1. Standardized path coefficients from study variables to infants' attention to frustration.
Note. Also depicted are paths from birth mother (BM) externalizing behaviors and explained variance in outcome above and below the mean on the adoptive mother (AM) affective state variable ("AM = high" and "AM = low," respectively).
$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Decomposition of Interaction Between BM Externalizing Behavior and AM Affective State

To examine the nature of this interaction, the model was run within the subsamples that were above and below the mean on the AM affective state variable; nonsignificant covariates (AF affective state, infant sex, and adoption openness) were omitted given the reduced subsample sizes. In the subsample above the mean on the AM affective state variable ($n = 184$), the model provided a good fit to the data, $\chi^2 = 5.17$, $df = 6$, $p = .523$; CFI = 1.00; RMSEA = .00. The path from BM externalizing problems to the infant attention to frustration was significant ($\beta = .35$, $p < .01$). This model was then rerun with the subsample below the mean on the AM affective state variable ($n = 164$). This model also fit the data well, $\chi^2 = 6.45$, $df = 6$, $p = .38$; CFI = .99; RMSEA = .02. The path from BM externalizing problems to infant attention was not significant ($\beta = -.01$).

The interaction decomposition models were rerun to examine whether the pattern of effects remained when the sample was split at 1 *SD* above and below the mean and at 0.5 *SD* above and below the mean on the AM affective state variable. Due to the reduced sample sizes, these models were considered preliminary. The standardized path coefficients from BM externalizing to infant attention to frustration were in the same direction as the mean-split analyses for both models: .34 at 1 *SD* above the mean of AM affective state ($n = 62$); .36 at 0.5 *SD* above the mean ($n = 121$, $p < .05$); −.04 at 1 *SD* below the mean ($n = 68$); and −.02 at 0.5 *SD* below the mean ($n = 121$). These follow-up analyses support the effects found with the mean-split approach. Infant attention to frustration was regressed on BM externalizing problem factor scores for the two mean-split groups (Figure 2), reflecting the linear regression results using observed (vs. latent) variables and listwise deletion of cases with missing data.

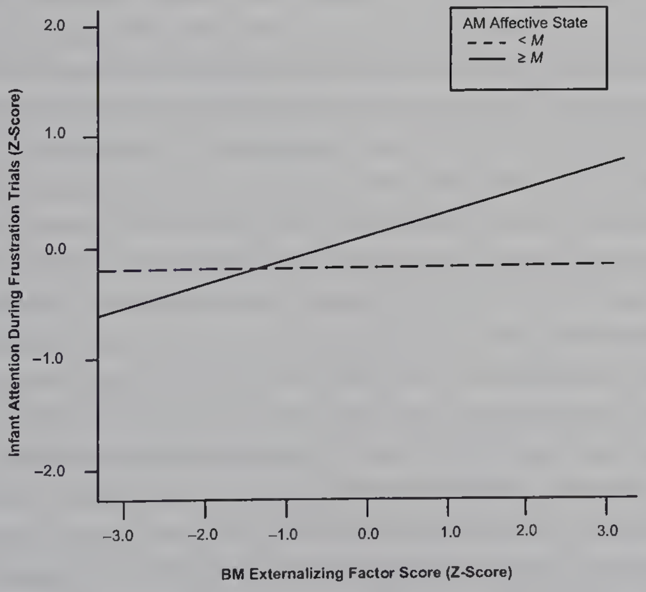


Figure 2. Relations between birth mother (BM) externalizing problems and infants' attention to frustration above and below the mean on the measure of adoptive mother (AM) affective state.

Finally, we ran a similar series of models to evaluate whether AF affective state interacted with BM externalizing to predict infant attention during the frustration task. In these models, AM affective state and the other covariates (described earlier) were controlled. All main effect and interaction paths were nonsignificant, suggesting that the effects described earlier are specific to AM affective state.

Linear Regression: Evaluation of BF × Adoptive Parent Effects

The G × E hypotheses were examined in a parallel way with BF data using linear regression techniques due to the nonconvergence of the latent factor and the smaller sample size. Externalizing behavior problems were represented by the factor score composed of BF ATOD dependence, delinquency, and novelty seeking indicators. Infant attention during the frustration trials was not predicted by this BF composite alone or in interaction with the AM affective state variable, $F(3, 81) = .587$, *ns*, or the AF affective state variable, $F(3, 78) = .255$, *ns*.

Exploratory Models Controlling for BM Prenatal ATOD Use

Next, exploratory models were run to test whether the effect of the interaction between BM externalizing problems and AM affective state on infants' attention during the frustration trials was better accounted for by BM prenatal ATOD use. In general, all models and path coefficients were regarded with caution in these analyses due to the functional relation between general and prenatal ATOD use (i.e., mothers who report never using will also report never using during pregnancy, and mothers who use during pregnancy typically use outside of pregnancy).

Birth mother prenatal ATOD use was entered in the full model as a mediator of interaction effects. That is, BM prenatal ATOD use was regressed on BM externalizing problems, AM affective state, and their interaction, and infant attention to frustration was regressed on BM prenatal ATOD use, attention during neutral trials, infant sex, AF affective state, and adoption openness but not directly on terms involved in the interaction. BM externalizing was significantly associated with BM prenatal ATOD use, but the interaction term was not significant. Additionally, BM prenatal ATOD use did not significantly predict infant attention to frustration. This model was compared with a model that allowed direct predictive paths from BM externalizing, AM affective state, and their interaction to infant attention during the frustration trials. The latter model fit the data better, Satorra–Bentler χ^2 difference ($df = 3$) = 15.46, $p < .05$, and the path from the interaction term to the infant attention outcomes was significant ($p < .01$). Similarly, within the subsample that was above the mean on AM affective state, the reduced model (i.e., exclusion of infant sex, AF affective state, and openness in adoption) provided a marginal fit to the data, $\chi^2 = 14.924$, $df = 9$, $p = .093$; CFI = .960; RMSEA = .060. BM externalizing problems predicted BM prenatal ATOD use ($\beta = .72$, $p < .001$) and infant attention during the frustration trials ($\beta = .48$, $p < .05$), whereas BM prenatal ATOD use did not predict infant attention during the frustration trials ($\beta = -.19$, $p = ns$). The model that fixed the direct path from BM externalizing problems to infant attention during the frustration trials did not fit the data adequately, $\chi^2 = 25.381$, $df = 10$, $p = .005$; CFI = .896; RMSEA = .091. Therefore, the findings did not support BM prenatal ATOD use as an explanation for the G × E findings described above.

Discussion

Recently, researchers from developmental psychology and behavioral genetic perspectives have called for and highlighted the potential value of integrating behavioral genetic and developmental research to improve outcomes for youth and families (Reiss & Leve, 2007; Rutter, 2005). In the current study, we made a preliminary effort to leverage the unique opportunity provided by an adoption design to understand how a specific family environmental experience—infant exposure to adoptive parent anxious and depressive symptoms—might moderate a genetic propensity for externalizing behavior very early in development and, as such, increase the understanding of causal mechanisms during early childhood.

The results suggest that infants are more likely to show heightened attention during a frustration task when they are exposed to an early rearing environment involving elevated maternal anxious and depressive symptoms. Although the direct association between maternal anxious and depressive symptoms and infant attention to frustration was modest in this study, this main effect finding is consistent with prior findings that demonstrate the significant effects of maternal depression in early childhood on young children's problem behaviors (Downey & Coyne, 1990). However, it is unique in that the parents and children were genetically unrelated in the present study, thereby suggesting a pure environmental mechanism of transmission that is unclouded by genetic factors. Beyond the direct effects of AM affective state, the SEM results suggest that maternal affective state moderates the effects of genetic risk for externalizing behavior. That is, the direct effect of birth parent externalizing behavior on infant attention to frustration was amplified in the context of elevated levels of AMs' anxious and depressive symptoms. Researchers have examined the moderating role of *genetic* risk for psychopathology (e.g., birth parent psychopathology; Riggins-Caspers, Cadoret, Knutson, & Langbehn, 2001), but, to our knowledge, this study is the first attempt to examine the affective state of rearing parents on moderating genetic influences in young children. As is shown in Figure 2, genetic risk for externalizing behavior was unrelated to infant attention when AMs were below the mean on anxious and depressive symptoms. Conversely, when AMs were above the mean, genetic risk was exacerbated, with the association between BM externalizing behavior and infant attention during a frustration task becoming significant ($\beta = .35$, $p < .01$).

The specific phenotype measured in the present study was infants' attention, interest, and persistence to a frustrating task. As such, this behavior was considered to reflect an inability to direct attention away from a stressful situation (rather than reflecting more general attention processes). The follow-up analyses examining whether the same G × E processes operated on infant attention during the neutral trials yielded nonsignificant results, suggesting the specificity of maternal affect in moderating effects to situations where the infant might feel frustrated and lack autonomy. Through a combination of genetic liability and emotionally unavailable and dysregulated maternal caregivers, these infants might begin to react with increased attention and intensity under conditions of frustration, perhaps in an effort to solicit caregiver attention or involvement from an anxious or depressed mother. Such reactions during infancy might be the forbearers of noncompliance and reactivity during toddlerhood, when children are in similarly frustrating situations and are subjected to constraints placed upon them by their caregivers (e.g., when a parent asks a child to clean up his or her toys). The reactions of some children can increase under such conditions, escalating into coercive cycles with the parent during childhood (Shaw et al., 1994; Shaw et al., 1998) and evolving into delinquency and antisocial behavior that continues into adolescence (Patterson, Reid, & Dishion, 1992).

Parental Sex Differences

Although support was found for direct and moderating effects of AM affective state on infant attention during a frustration task, the AF results did not support the hypotheses. Similarly, the G × E interaction was not significant when the BF data were modeled. Together, these nonsignificant results for fathers suggest a pattern of maternal linkages to externalizing behavior in young children, in terms of both the rearing environment and genetic influences. The lack of significant findings for the unique contributions of AF affective state was not surprising; prior work has found stronger associations between maternal affective state and child adjustment than between paternal affective state and child adjustment (Meadows et al., 2007). Measuring the respective caregiving roles of both parents in future studies would help elucidate the extent to which this rearing parent environmental effect results from differential levels of caregiving between mothers and fathers or whether there is something unique about maternal affective state irrespective of the level of her caregiving. Alternatively, given that the parenting factors used in the present study have been validated primarily with samples of mothers, other aspects of paternal caregiving in early or later development (e.g., facilitating problem solving in middle childhood) might play a more important role in the development of child problem behavior.

The lack of significant BF effects was unexpected and suggests the possibility of sex-linked transmission of genetic risk for externalizing behavior. Although this is the largest BF sample in a single prospective adoption design collected to date, we could not test models that aggregate genetic risk by forming a composite between BM and BF data because of the large number of nonparticipating BFs and the lack of convergence of indicators for the participating BFs. Further, the results from the BF regression analyses indicate that the direction of effects does not resemble that of BMs, suggesting that a larger sample would not alter the pattern of findings. Sex differences in the heritability of externalizing behavior have been examined in a number of studies. Although some researchers have reported similar patterns of heritability across the sexes (Rhee & Waldman, 2002), others have found that genetic influences are significantly higher in females than in males. For example, Tuvblad, Grann, and Lichtenstein (2006) found heritability estimates of 6% for adolescent boys' antisocial behavior and 59% for adolescent girls' antisocial behavior.

A polygenetic multiple threshold model can be used as a framework to consider the sex differences found in this study (Rhee & Waldman, 2002). In such a model, the sex less afflicted with a behavior might require a greater liability to express that behavior. In the case of externalizing behaviors, females are the less affected sex; thus, the relatives of affected females might be at greater risk of developing externalizing problems than relatives of affected males. In the present analyses, the associations between BM externalizing behavior (compared to BF externalizing behavior) and infant attention during a frustration task were more pronounced, thereby permitting sufficient variation to allow for the detection of environmental moderation of this genetic effect. In addition, as is shown in Table 1, the mean levels of externalizing behavior were only modestly lower for BMs than for BFs, an atypical finding for adult samples that provides some corroborating evidence for the polygenetic multiple threshold model. Parallel evidence in support of polygenetic multiple threshold processes comes from clinical research involving a gender paradox in the severity of disorders. For disorders with an unequal sex ratio, individuals with the lower prevalence rates tend to be more seriously affected and to have more co-occurring diagnoses (Loeber & Keenan, 1994). For example, the likelihood of girls with attention deficit disorder to develop conduct disorder is 40 times greater than for girls without attention deficit disorder, compared to a factor of 14.7 for boys (Szatmari, Boyle, & Offord, 1989). Although the lack of significant effects for the BF analyses should be interpreted with caution because they were counter to study predictions, the two lines of research described above suggest that future research should consider sex-linked explanations when examining G × E interaction processes on externalizing behaviors in early childhood.

Future Directions: Implications for the Development of Preventive Interventions

In conducting this study, we had an implicit goal of exploring possibilities for the development of future preventive interventions by capitalizing on the EGDS adoption study design, which provides information about the processes whereby early environmental experiences can offset or exacerbate genetic strengths and liabilities. As has been noted by others, adoption is an effective intervention, particularly for children living in suboptimal environments who are placed in more enriched environments within the first few months of life (Rutter, O'Connor, & English and Romanian Adoptees Study Team, 2004; van IJzendoorn & Juffer, 2006). Children adopted at an early age typically fare better than their nonadopted counterparts on a range of cognitive and social outcomes. However, in addition to mean level shifts in behavior resulting from adoption, the full adoption design permits the examination of the mechanisms whereby differing early environmental experiences can lead to healthy or less optimal adjustment. In the present study, AM anxious and depressive symptoms moderated the effect of genetic risk for early externalizing problems. If this effect were to be replicated and extended in future research, preventive interventions targeting mothers with affective problems might improve the mental health of the caregiver, in turn alleviating exposure to adverse environmental conditions that might offset the expression of children's genetic risk for externalizing behavior. Although this possibility cannot be directly addressed in the present study until the link between infant frustration and later externalizing problems is confirmed, it is tentatively proposed as a hypothesis for further investigation. Notwithstanding this caveat, the results from two recent intervention trials have provided the first evidence that interventions aimed at reducing parental depression might significantly reduce child behavior problems (Shaw, Dishion, Connell, Wilson, & Gardner, 2009; Weissman et al., 2006), thereby supporting this hypothesis.

Alternative Explanations

Although the results of the present analyses are unique and appear to be robust, several alternative explanations for AM environmental moderation of genetic effects could be possible. First, intrauterine exposure to toxins might mask or confound genetic effects. Although BM prenatal ATOD use does not appear to be a causal mechanism in the associations found here, additional prenatal experiences (e.g., exposure to environmental toxins) have not been examined; thus, other prenatal factors could be associated with the pattern of findings described here.

Second, trends in adoption practices such as selective placement (agency matching of birth and adoptive parent characteristics) and openness (contact and knowledge between birth and adoptive families) can pose a threat to the assumptions underlying the adoption design and bias model estimates. For example, an adopted child might be more likely to resemble his or her birth parents (artificially inflating genetic estimates) if the birth parents are in direct contact with the child or if the birth parents have actively selected an adoptive home that closely mirrors their own environment. As such, an openness construct was included in the analytical models; this construct neither yielded a significant association with the dependent measure nor accounted for significant variance in the relation between BM externalizing behavior, AM affective state, and infant attention during a frustration task (Figure 1). Ruling out the possible effects of selective placement is more complex because the associations between a birth parent characteristic and an adoptive parent characteristic could indicate selective placement or could indicate genetically influenced evocative behaviors in the child that have modified adoptive parent behavior to be more similar to birth parent characteristics. Previously, we correlated a set of birth parent characteristics with a set of adoptive family characteristics that were *unlikely* to be influenced by evocative effects (i.e., demographic characteristics) to consider the possible effects of selective placement (Leve et al., 2007). No relation was significant. Thus, systematic selective placement was not identified, although it cannot be fully ruled out as a possible explanation for the present effects.

Third, the G × E effects found in the present study could have resulted from evocative genotype–environment correlation (*r*GE), which has been found in other samples of infants and toddlers (Forget-Dubois et al., 2006). In the present study, evocative *r*GE would have been noted if the genetically influenced externalizing behavior of the infant directly influenced AM anxious and depressive symptoms and confounded the G × E effects. Although longitudinal designs are needed to truly separate the directionality of evocative effects, the correlation between AM affective state and BM externalizing behavior in this study was nonsignificant. Similarly, the correlations between fathers (adoptive and birth) were nonsignificant (Table 1). These results suggest that *r*GE is likely not confounding the moderating effects shown by AM affective state.

Fourth, in the present study, we examined a single outcome. Based on the research reviewed earlier, we hypothesized that attention, interest, and persistence toward a frustrating situation would reflect an underlying genetic predisposition toward later externalizing behavior, with children at genetic risk showing heightened attention when frustrating constraints were placed upon them. Alternatively, the barrier frustration task might simply elicit vulnerabilities to negative affectivity rather than differences in attention processes. To test this alternative hypothesis, we examined the final models using negative affect during the barrier task as the outcome using the same coding system as the attention variable. No significant G × E effects were found. Further, there were significantly stronger associations between negative affect and attention in the control trials ($r = -.53$, $p < .001$) than in the frustration trials ($r = -.15$, $p < .05$), suggesting the relative independence of attention and negative affect within the context of a frustrating situation.

Fifth, based on prior findings in the literature, we hypothesized that the primary genetic risk factor associated with infant attention would be externalizing problems and that the primary early environmental risk factor would be adoptive parent anxious and depressive symptoms. However, given the modest amount of variance accounted for in our results, it is plausible that additional genetic influences and family environmental contexts affect infant attention to frustration. Although rates of adoptive parent externalizing behavior were low in this sample, we tested whether similar main effects and interaction effects would be found when adoptive parent antisocial behavior or novelty seeking was entered as the environmental measure (using identical measures as were used for birth parents). In this post hoc set of analyses, no main effects or interaction effects were found, providing converging evidence for maternal affective state as a central environmental variable from birth to 9 months. We next examined alternative genetic risk factors and conducted a post hoc analysis of whether BM anxiety and depression (using identical measures as were used for adoptive parents) would show associations with infant attention during a frustration task. Although there was no main effect, the interaction between BM affective state and AM affective state was significant, suggesting a potential explanatory role of BM affective state. However, given the relatively high association between BM externalizing problems and BM affective state in this sample ($r = .40$, $p < .001$) and the fact that the original interaction between BM externalizing problems and AM affective state remained significant while controlling for BM affective state, the effects of BM affective state do not appear to be as strong as the effects of BM externalizing problems in this study. Further, the findings from prior research on attention to frustration have suggested associations with later externalizing problems (rather than internalizing problems), suggesting that our reliance on prior research to guide the selection of genetic and environmental risk variables in the present study was appropriate.

Limitations

To our knowledge, the present study is the first to examine how early exposure to maternal anxious and depressive symptom moderates genetic influences on risk for later externalizing behavior. The use of a prospective adoption design enabled us to make associations between the phenotypic behavior of the birth parent and the phenotypic behavior of an infant to examine underlying inherited predispositions toward externalizing behavior. However, the present study is not without limitations. First, because cross-sectional data were used, longitudinal pathways could not be tested. The longitudinal nature of the study will enable us to examine, in subsequent years, how the moderating effects of AM affective state unfold over time. Of relevance to informing preventive intervention, it will be critical to examine whether the mechanisms affecting infant attention are in fact found to lead the child toward a trajectory of early and persistent externalizing problems. Without such longitudinal data, we can only speculate about the intermediary phenotypic pathways to externalizing behavior and cannot make conclusive interpretations. Further, the moderating effects of the early home environment were already present at 9 months of age; studies beginning earlier in development are needed to help define the causal mechanisms whereby early exposure to maternal anxiety and depression result in the expression of risk in the child.

In addition, although findings from prior research on attention to frustration have suggested associations with later externalizing problems, it is likely that additional genetic influences and family environmental contexts affect infant attention to frustration. Specifically, attachment is a key environmental influence on later infant externalizing behavior (Smeekens et al., 2007), and cognitive functioning is known to be a heritable characteristic related to externalizing behavior (DeYoung et al., 2006). Such variables (not assessed at our 9-month assessment) might be associated with the infant attention to frustration variable. Additional research is needed to test these alternate environmental and genetic pathways.

Finally, although our sample was representative of birth and adoptive families from the agencies participating in this study, the adoptive families had high educational and economic backgrounds and low rates of externalizing problems compared to national norms. Thus, the discrepancy in sociodemographic characteristics often found between birth and adoptive parents was replicated in the present study, indicating the potential restriction of range in the environment that has been noted in prior reviews of the adoption design (Stoolmiller, 1999). However, in a recent systematic test of range restriction biases, McGue et al. (2007) found negligible effects on estimates of heritability and the environment, even when range restriction was present. Further, the adoptive parents in our study had an adequate range of depression and anxiety symptoms (Leve et al., 2007). Other researchers have found that genetic influences on externalizing behavior are more important in families with more advantaged environments, whereas the shared environment is more important in more economically disadvantaged environments (Tuvblad et al., 2006). Thus, the generalizability of findings to high-risk home environments and more ethnically diverse samples needs to be documented before stronger conclusions can be drawn.

References

Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumaker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 243–277). Mahwah, NJ: Erlbaum.

Beck, A., & Steer, R. (1993a). *Beck Anxiety Inventory manual*. San Antonio, TX: Psychological Corporation.

Beck, A., & Steer, R. (1993b). *Beck Depression Inventory manual*. San Antonio, TX: Psychological Corporation.

Blair, C. (2002). Early intervention for low birth weight, preterm infants: The role of negative emotionality in the specification of effects. *Development and Psychopathology, 14*, 311–332.

Campbell, S. B., Matestic, P., von Stauffenberg, C., Mohan, R., & Kirchner, T. (2007). Trajectories of maternal depressive symptoms, maternal sensitivity, and children's functioning at school entry. *Developmental Psychology, 43*, 1202–1215.

Caspi, A., Henry, B., McGee, R. O., Moffitt, T. E., & Silva, P. A. (1995). Temperamental origins of child and adolescent behavior problems: From age three to fifteen. *Child Development, 66*, 55–68.

Caspi, A., McClay, J., Moffitt, T. E., Mill, J., Martin, J., Craig, I. W., et al. (2002). Role of genotype in the cycle of violence in maltreated children. *Science, 297*, 851–854.

Caspi, A., Moffitt, T. E., Newman, D. L., & Silva, P. A. (1996). Behavioral observations at age 3 years predict adult psychiatric disorders. *Archives of General Psychiatry, 53*, 1033–1039.

Caspi, A., Moffitt, T. E., Thornton, A., Freedman, D., Arnell, J. W., Harrington, H., et al. (1996). The life history calendar: A research and clinical assessment method for collecting retrospective event-history data. *International Journal of Methods in Psychiatric Research, 6*, 101–114.

Cicchetti, D., Rogosch, F. A., & Toth, S. L. (1998). Maternal depressive disorder and contextual risk: Contributions to the development of attachment insecurity and behavior problems in toddlerhood. *Development and Psychopathology, 10*, 283–300.

Cloninger, C. R., Svaric, D. M., & Przybeck, T. R. (1993). A psychobiological model of temperament and character. *Archives of General Psychiatry, 50*, 975–990.

Crockenberg, S. C., Leerkes, E. M., & Barrig Jo, P. S. (2008). Predicting aggressive behavior in the third year from infant reactivity and regulation as moderated by maternal behavior. *Development and Psychopathology, 20*, 37–54.

DeYoung, C. G., Peterson, J. B., Seguin, J. R., Mejia, J. M., Pihl, R. O., Beitchman, J. H., et al. (2006). The dopamine D4 receptor gene and moderation of the association between externalizing behavior and IQ. *Archives of General Psychiatry, 63*, 1410–1416.

Dogan, S., Lei, A., Milne-Kahn, J., Pong, H., Wu, E., & Conger, R. (2005). *Family interaction behavioral codes for the Early Growth and Development Study*. Unpublished coding manual.

Downey, G., & Coyne, J. C. (1990). Children of depressed parents: An integrative review. *Psychological Bulletin, 108*, 50–76.

Elliott, D. S., Huizinga, D., & Ageton, S. S. (1985). *Explaining delinquency and drug use*. Beverly Hills, CA: Sage.

Feinberg, M. E., Button, T. M. M., Neiderhiser, J. M., Reiss, D., & Hetherington, M. (2007). Parenting and adolescent antisocial behavior and depression: Evidence of Genotype × Parenting Environment interaction. *Archives of General Psychiatry, 64*, 457–465.

Fisher, P. A., Stoolmiller, M., Gunnar, M. R., & Burraston, B. (2007). Effects of a therapeutic intervention for foster preschoolers on diurnal cortisol activity. *Psychoneuroendocrinology, 32*, 892–905.

Forget-Dubois, N., Boivin, M., Dionne, G., Pierce, T., Tremblay, R. E., & Pérusse, D. (2006). A longitudinal twin study of the genetic and environmental etiology of maternal hostile-reactive behavior during infancy and toddlerhood. *Infant Behavior and Development, 30*, 453–465.

Ge, X., Natsuaki, M. N., Martin, D., Leve, L. D., Neiderhiser, J. M., Shaw, D. S., et al. (2008). Bridging the divide: Openness in adoption and post-adoption psychosocial adjustment among birth and adoptive parents. *Journal of Family Psychology, 22*, 529–540.

Goldsmith, H. H., & Rothbart, M. K. (1996). *The Laboratory Temperament Assessment Battery (LAB-TAB) Prelocomotor version*. Unpublished manual.

Hershberger, S. L. (1994). Genotype-environment interaction and correlation. In J. C. DeFries, R. Plomin, & D. W. Fulker (Eds.), *Nature and nurture during middle childhood* (pp. 281–294). Malden, MA: Blackwell.

Hiroi, N., & Agatsuma, S. (2005). Genetic susceptibility to substance dependence. *Molecular Psychiatry, 10*, 336–344.

Johnson, M. H., Posner, M. I., & Rothbart, M. K. (1991). Components of visual orienting in early infancy: Contingency learning, anticipatory looking, and disengaging. *Journal of Cognitive Neuroscience, 3*, 335–344.

Kessler, R., Andrews, G., Mroczek, D., Ustun, B., & Wittchen, H. (1998). The World Health Organization Composite International Diagnostic Interview Short-Form (CIDI–SF). *International Journal of Methods in Psychiatric Research, 7*, 171–184.

Laucht, M., Becker, K., El-Faddagh, M., Hohm, E., & Schmidt, M. H. (2005). Association of the DRD4 exon III polymorphism with smoking in 15-year-olds: A mediating role for novelty seeking? *Journal of the American Academy of Child and Adolescent Psychiatry, 44*, 477–484.

Laucht, M., Becker, K., & Schmidt, M. H. (2006). Visual exploratory behavior in infancy and novelty seeking in adolescence: Two developmentally specific phenotypes of DRD4? *Journal of Child Psychology and Psychiatry, 47*, 1143–1151.

Laucht, M., Skowronek, M. H., Becker, K., Schmidt, M. H., Esser, G., Schulze, T. G., et al. (2007). Interacting effects of the dopamine transporter gene and psychosocial adversity on attention-deficit/hyperactivity disorder symptoms among 15-year-olds from a high-risk community sample. *Archives of General Psychiatry, 64*, 585–590.

Leve, L. D., Neiderhiser, J. M., Ge, X., Scaramella, L. V., Conger, R. D., Reid, J. B., et al. (2007). The Early Growth and Development Study: A prospective adoption design. *Twin Research and Human Genetics, 10*, 84–95.

Leve, L. D., Neiderhiser, J. M., Scaramella, L. V., & Reiss, D. (2008). The Early Growth and Development Study: Using the prospective adoption design to examine genotype–environment interplay. *Acta Psychologica Sinica, 40*, 1106–1115.

Loeber, R., & Keenan, K. (1994). The interaction between conduct disorder and its comorbid conditions: Effects of age and gender. *Clinical Psychology Review, 14*, 497–523.

Lyons-Ruth, K., Easterbrooks, M. A., & Cibelli, C. D. (1997). Infant attachment strategies, infant mental lag, and maternal depressive symptoms: Predictors of internalizing and externalizing problems at age 7. *Developmental Psychology, 33*, 681–692.

Martin, J. (1981). A longitudinal study of the consequences of early mother–infant interaction: A microanalytic approach. *Monographs of the Society for Research in Child Development, 46*, 1–58.

McGue, M., Keyes, M., Sharma, A., Elkins, I., Legrand, L., Johnson, W., et al. (2007). The environments of adopted and non-adopted youth: Evidence on range restriction from the Sibling Interaction Behavior Study (SIBS). *Behavior Genetics, 37*, 449–462.

Meadows, S. O., McLanahan, S. S., & Brooks-Gunn, J. (2007). Parental depression and anxiety and early childhood behavior problems across family types. *Journal of Marriage and Family, 69*, 1162–1177.

Miles, D. R., & Carey, G. (1997). Genetic and environmental architecture of human aggression. *Journal of Personality and Social Psychology, 7*, 207–217.

Moffitt, T. E., Caspi, A., & Rutter, M. (2005). Strategy for investigating interactions between measured genes and measured environments. *Archives of General Psychiatry, 62*, 473–481.

Muthén, L. K., & Muthén, B. O. (2006). *Mplus user's guide* (4th ed.). Los Angeles: Author.

O'Connor, T. G., Caspi, A., DeFries, J. C., & Plomin, R. (2003). Genotype-environment interaction in children's adjustment to parental separation. *Journal of Child Psychology and Psychiatry, 44*, 849–856.

Olds, D. (2002). Prenatal and infancy home visiting by nurses: From randomized trials to community replication. *Prevention Science, 3*, 153–172.

Patterson, G. R., Reid, J. B., & Dishion, T. J. (1992). *A social learning approach: IV. Antisocial boys*. Eugene, OR: Castalia.

Plomin, R., DeFries, J. C., & Fulker, D. W. (Eds.). (1988). *Nature and nurture during infancy and early childhood*. New York: Cambridge University Press.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience, 13*, 25–42.

Reiss, D., & Leve, L. D. (2007). Genetic expression outside the skin: Clues to mechanisms of Genotype × Environment interaction. *Development and Psychopathology, 19*, 1005–1027.

Reiss, D., Neiderhiser, J., Hetherington, E. M., & Plomin, R. (2000). *The relationship code: Deciphering genetic and social patterns in adolescent development*. Cambridge, MA: Harvard University Press.

Rhee, S. H., & Waldman, I. D. (2002). Genetic and environmental influences on antisocial behavior: A meta-analysis of twin and adoption studies. *Psychological Bulletin, 128*, 490–529.

Riggins-Caspers, K. M., Cadoret, R. J., Knutson, J. F., & Langbehn, D. (2001). Biology-environment interaction and evocative biology-environment correlation: Contributions of harsh discipline and parental psychopathology to problem adolescent behaviors. *Behavior Genetics, 33*, 205–220.

Rutter, M. (2005). Environmentally mediated risks for psychopathology: Research strategies and findings. *Journal of the American Academy of Child and Adolescent Psychiatry, 44*, 3–18.

Rutter, M. (2006). *Genes and behavior: Nature-nurture interplay explained*. Oxford, UK: Blackwell.

Rutter, M., O'Connor, T. G., & English and Romanian Adoptees Study Team. (2004). Are there biological programming effects for psychological development? Findings from a study of Romanian adoptees. *Developmental Psychology, 40*, 81–94.

Shackman, J. E., Shackman, A. J., & Pollak, S. D. (2007). Physical abuse amplifies attention to threat and increases anxiety in children. *Emotion, 7*, 838–852.

Shaw, D. S., Connell, A., Dishon, T. J., Wilson, M. N., & Gardner, F. (2009). Improvements in maternal depression as a mediator of intervention effects on early childhood problem behavior. *Development and Psychopathology, 21*, 417–439.

Shaw, D. S., Dishion, T. J., Supplee, L., Gardner, F., & Arnds, K. (2006). A family-centered approach to the prevention of early-onset antisocial behavior: Two-year effects of the Family Check-Up in early childhood. *Journal of Consulting and Clinical Psychology, 74*, 1–9.

Shaw, D. S., Keenan, K., & Vondra, J. l. (1994). Developmental precursors of externalizing behavior: Ages 1 to 3. *Developmental Psychology, 30*, 355–364.

Shaw, D. S., Winslow, E. B., Owens, E. B., Vondra, J. I., Cohn, J. F., & Bell, R. Q. (1998). The development of early externalizing problems among children from low-income families: A transformational perspective. *Journal of Abnormal Child Psychology, 26*, 95–107.

Smeekens, S., Riksen-Walraven, J. M., & van Bakel, H. J. A. (2007). Multiple determinants of externalizing behavior in 5-year-olds: A longitudinal model. *Journal of Abnormal Child Psychology, 35*, 347–361.

Stifter, C. A., & Braungart, J. M. (1995). The regulation of negative reactivity in infancy: Function and development. *Developmental Psychology, 31*, 448–455.

Stoolmiller, M. (1999). Implications of the restricted range of family environments for estimates of heritability and nonshared environment in behavior-genetic adoption studies. *Psychological Bulletin, 125*, 392–409.

Szatmari, P., Boyle, M., & Offord, D. R. (1989). ADHD and CD: Degree of diagnostic overlap and difference among correlates. *Journal of the American Academy of Child and Adolescent Psychiatry, 28*, 865–872.

Thapar, A., O'Donovan, M., & Owen, M. J. (2005). The genetics of attention deficit hyperactivity disorder. *Human Molecular Genetics, 14*, R275–R282.

Tuvblad, C., Grann, M., & Lichtenstein, P. (2006). Heritability for adolescent antisocial behavior differs with socioeconomic status: Gene-environment interaction. *Journal of Child Psychology and Psychiatry, 47*, 734–743.

van IJzendoorn, M. H., & Juffer, F. (2006). Adoption as intervention. Meta-analytic evidence for massive catch-up and plasticity in physical, socio-emotional, and cognitive development. *Journal of Child Psychology and Psychiatry, 47*, 1228–1245.

Weissman, M. M., Pilowsky, D. J., Wickramaratne, P. J., Talati, A., Wisniewski, S. R., Fava, M., et al. (2006). Remissions in maternal depression and child psychopathology: A STAR*D-Child Report. *Journal of the American Medical Association, 295*, 1389–1398.

Wothke, W. (2000). Longitudinal and multi-group modeling with missing data. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multiple group data: Practical issues, applied approaches and specific examples* (pp. 269–281). Mahwah, NJ: Erlbaum.

Child Development, January/February 2010, Volume 81, Number 1, Pages 357–367

Building a New Biodevelopmental Framework to Guide the Future of Early Childhood Policy

Jack P. Shonkoff
Harvard University

Four decades of early childhood policy and program development indicate that evidence-based interventions can improve life outcomes, and dramatic advances in the biological and behavioral sciences now provide an opportunity to augment those impacts. The challenge of reducing the gap between what we know and what we do to promote the healthy development of young children is to view current best practices as a starting point and to leverage scientific concepts to inspire fresh thinking. This article offers an integrated, biodevelopmental framework to promote greater understanding of the antecedents and causal pathways that lead to disparities in health, learning, and behavior in order to inform the development of enhanced theories of change to drive innovation in policies and programs.

Dramatic advances in neuroscience, molecular biology, genomics, and the behavioral and social sciences are deepening our understanding of how healthy development happens, how it can be derailed, and what societies can do to keep it on track. We now know that genes provide the initial blueprint for building brain architecture, environmental influences affect how the neural circuitry actually gets wired, and reciprocal interactions among genetic predispositions and early experiences affect the extent to which the foundations of learning, behavior, and both physical and mental health will be strong or weak (Fox, Levitt, & Nelson, 2010; Meaney, 2010). A strong foundation lays the groundwork for responsible citizenship, economic prosperity, healthy communities, and successful parenting of the next generation. A weak foundation can seriously undermine the social and economic vitality of a nation (Knudsen, Heckman, Cameron, & Shonkoff, 2006).

Decades of research in child development have taught us that families and communities play the central role and bear most of the costs of providing the supportive relationships and positive learning experiences that young children need for healthy development (Shonkoff & Phillips, 2000). There also has been considerable documentation of the extent to which public policies and formal services can enhance developmental outcomes for young children living in a wide variety of circumstances (Karoly, Kilburn, & Cannon, 2005). Building on this extensive knowledge base, striking new discoveries in the biological sciences are presenting policy makers and civic leaders with provocative insights into the far-reaching influences of early developmental processes that were not appreciated as recently as a decade ago. The extent to which advances in molecular biology are already beginning to transform the way we diagnose and treat disease illustrates one example of this new world. The challenge before us is to capitalize on the capacity of this scientific revolution to stimulate creative new ways of thinking about how to address a much broader range of societal concerns, including education reform, workforce development, health promotion, prevention of disease and disability, protection of children from the consequences of maltreatment and exploitation, reduction in violent crime, and alleviation of poverty.

The author acknowledges the enormous contributions of past and present members of the National Scientific Council on the Developing Child, including W. Thomas Boyce, Judy Cameron, Greg Duncan, Nathan Fox, William Greenough, Megan Gunnar, Eric Knudsen, Pat Levitt, Betsy Lozoff, Linda Mayes, Bruce McEwen, Charles Nelson, Deborah Phillips, and Ross Thompson; members of the National Forum on Early Childhood Program Evaluation, including Jeanne Brooks-Gunn, Greg Duncan, Bernard Guyer, Katherine Magnuson, Deborah Phillips, Helen Raikes, and Hirokazu Yoshikawa; associates of the FrameWorks Institute led by Susan Nall Bales; and staff of the Center on the Developing Child at Harvard University, led by Gillian Najarian and Al Race.

Correspondence concerning this article should be addressed to Jack P. Shonkoff, Center on the Developing Child, Harvard University , 50 Church Street, 4th Floor, Cambridge, MA 02138. Electronic mail may be sent to jack_shonkoff@harvard.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8101-0023

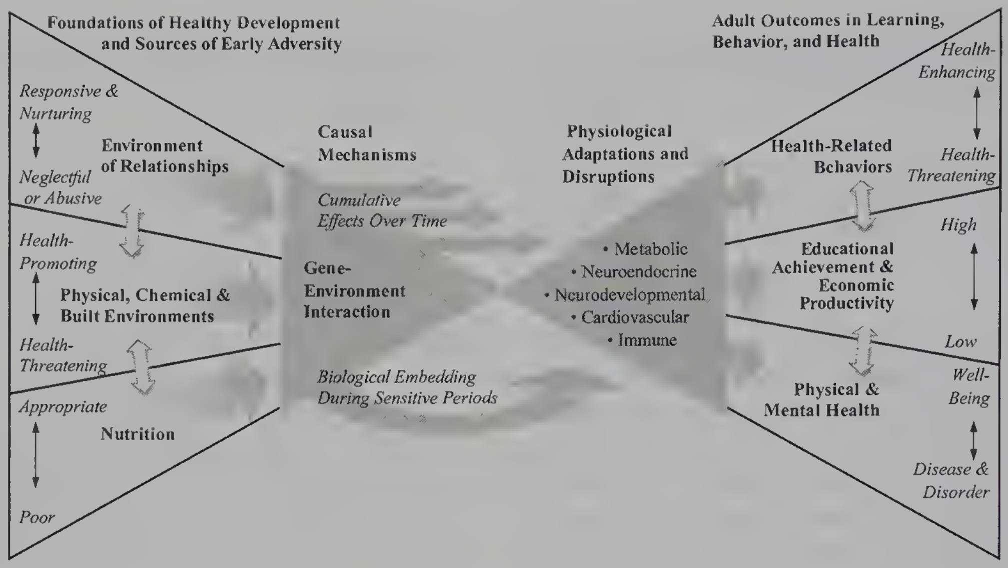

Foundations of Healthy Development and Sources of Early Adversity
Responsive & Nurturing
Environment of Relationships
Neglectful or Abusive
Health-Promoting
Physical, Chemical & Built Environments
Health-Threatening
Appropriate
Nutrition
Poor
Causal Mechanisms
Cumulative Effects Over Time
Gene-Environment Interaction
Biological Embedding During Sensitive Periods
Physiological Adaptations and Disruptions
• Metabolic
• Neuroendocrine
• Neurodevelopmental
• Cardiovascular
• Immune
Adult Outcomes in Learning, Behavior, and Health
Health-Enhancing
Health-Related Behaviors
Health-Threatening
High
Educational Achievement & Economic Productivity
Low
Well-Being
Physical & Mental Health
Disease & Disorder

Figure 1. A biodevelopmental framework for understanding the origins of disparities in learning, behavior, and health.

Developmental Frameworks and Theories of Change

Current Models

Early childhood policies and practices over the past several decades have been guided by several theoretical models of human development that have been refined over time. These include the transactional model formulated by Sameroff and Chandler (1975) and later adapted to the challenges of early childhood intervention by Sameroff and Fiese (1990, 2000), the ecological model articulated by Bronfenbrenner (1979), and the concepts of vulnerability and resilience developed by Werner and Smith (1982), Garmezy and Rutter (1983), and Rutter (2000). Together these models underscore the extent to which life outcomes are influenced by a dynamic interplay among the cumulative burden of risk factors and the buffering effects of protective factors that can be identified within the individual, family, community, and broader socioeconomic and cultural contexts. Each of these models also emphasizes the influence of reciprocal child–adult interactions in the developmental process, thereby underscoring the importance of stable and nurturing relationships, as well as the recognition that young children play an active role in their own development. The challenges of applying these multidimensional models, however, lie in their complexity. That is to say, the acknowledgment of numerous, interactive influences on developmental outcomes underscores both the futility of searching for simple solutions to complicated problems and the difficulty in choosing which variables to include and which to omit when one is designing a specific policy, program, or research project.

Building on these broadly supported models of child development, the field of early intervention for children living in disadvantaged circumstances (most typically associated with poverty) in the United States is currently driven by a theory of change that emphasizes the provision of enriched learning opportunities for the children and a combination of parenting education and support services for their families (usually focused exclusively on mothers), in community-based centers and/or the home. Over four decades of program development and evaluation, this basic model has been implemented successfully in a number of flagship demonstration projects, each of which has produced desirable impacts on a range of long-term outcomes, including increased rates of high school graduation and adult earnings, as well as reductions in special education referrals, welfare dependence, and incarceration, among others (Heckman, 2006; Shonkoff & Phillips, 2000).

Building an Enhanced Framework to Inform Innovative Policy

Although the basic principles of development described above have stood up well over time, advances in the biological and social sciences now offer an opportunity to formulate an augmented framework to drive a new generation of early childhood policies and practices. Support for this new framework is underscored by a rich and rapidly growing evidence base that illustrates the extent to which early experiences are biologically embedded in the development of multiple organ systems, with long-term impacts on metabolic regulation and cardiovascular health as well as the mastery of cognitive, language, and social skills (Hertzman, 2000; Meaney, 2010; Shonkoff, Boyce, & McEwen, 2009). Beyond its potential benefits for policies that affect all young children, the proposed biodevelopmental framework alerts decision makers to the particular importance of addressing the needs of those who are the most disadvantaged at the earliest ages. Equally important, yet potentially more challenging from a policy planning and implementation perspective, new research into individual differences in biological sensitivity to context offers significant promise in helping to explain why some children appear to do well in the face of adversity and why selected interventions appear to be effective for some children and not for others (Obradovic, Bush, Stamperdahl, Adler, & Boyce, 2010; O'Neal et al., 2010).

This new framework for research and policy is informed by a growing body of evidence that the foundations of healthy development and the origins of many impairments can be found among biological "memories" that are created through gene–environment interactions in the early years of life, in some cases beginning as early as the prenatal period (e.g., Barker, 1997; Davis & Sandman, 2010; D'Onofrio et al., 2010). These patterns, often mediated through epigenetic modifications, can be manifested within the circuitry of the developing brain and the physiological systems that influence a variety of biological functions such as stress management, immunological responsiveness, metabolic and neuroendocrine regulation, and cardiovascular integrity. The scientific concepts that explain these phenomena are derived from evolutionary biology, grounded in the assumption that the immature organism "reads" salient environmental characteristics in the service of developing the capacity to adapt to the environment in which it "expects" it will live. When early experiences are nurturing, contingent, stable, and predictable, healthy brain development is promoted, and other organ regulatory systems are facilitated. When early experiences are fraught with threat, uncertainty, neglect, or abuse, stress management systems are overactivated, and the consequences can include disruptions of developing brain circuitry as well as the establishment of a short fuse for subsequent activation of the stress response that leads to greater vulnerability to a host of chronic diseases. Although more research is needed to elucidate the precise nature and full range of causal mechanisms that mediate these effects, there is compelling evidence from recent studies of young children who have been abused or subjected to chronic neglect in institutionalized settings to support the conclusion that significant adversity early in life can induce physiological responses in the service of short-term survival benefits that come at considerable cost to long-term adaptive capacities as well as both physical and mental health (Cicchetti, Rogosch, Gunnar, & Toth, 2010; Pollak et al., 2010). The relation between low birth weight associated with severe nutritional deficiencies in utero and greater risk of subsequent metabolic syndrome, central body fat distribution, obesity, hypertension, insulin resistance, and cardiovascular disease is another example of this phenomenon (Barker, Osmond, Forsen, Kajantie, & Erikson, 2005).

In an effort to educate policy makers about the biology of adversity and its consequences for both health and development, the National Scientific Council on the Developing Child (2005) proposed a conceptually guided taxonomy based on three categories of stress experience—positive, tolerable, and toxic—to differentiate normative life challenges that are growth promoting from significant threats to long-term health and development that warrant intervention. It is important to note that these three categories refer to the physiological expression of the stress response and not to the specific stressors themselves. Although the underlying biology of these distinctions awaits empirical validation, their conceptual basis is grounded in well-established scientific principles.

Positive stress is characterized by moderate, short-lived increases in heart rate, blood pressure, serum glucose, and circulating levels of stress hormones such as cortisol and inflammatory cytokines such as interleukin-6. Precipitants include the challenges of dealing with frustration, adjusting to a new child-care setting, and other normative experiences. The essential characteristic of positive stress is that it is an important aspect of healthy development that is experienced in the context of stable and supportive relationships that facilitate adaptive responses that restore the stress response system to baseline.

Tolerable stress refers to a physiological state that could potentially disrupt brain architecture (e.g., through cortisol-induced damage of neural circuits in the hippocampus) but is buffered by supportive relationships that facilitate adaptive coping. Precipitants include the death or serious illness of a family member, parental divorce, homelessness, a natural disaster, or community violence. The defining characteristic of tolerable stress is the support provided by invested adults that helps restore the body's stress-response systems to baseline, thereby preventing neuronal disruptions that could lead to long-term consequences such as posttraumatic stress disorder.

Toxic stress refers to strong, frequent, and/or prolonged activation of the body's stress-response systems in the absence of the buffering protection of stable adult support. Major risk factors include extreme poverty, recurrent physical and/or emotional abuse, chronic neglect, severe maternal depression, parental substance abuse, and family violence. The defining characteristic of toxic stress is that it disrupts brain architecture, adversely affects other organs, and leads to stress management systems that establish relatively lower thresholds for responsiveness that persist throughout life, thereby increasing the risk of stress-related disease or disorder as well as cognitive impairment well into the adult years.

The basic elements of the biodevelopmental framework presented in this article are organized within three sets of target domains: (a) interactions among foundations of healthy development and sources of early adversity, (b) measures of physiological adaptation and disruption, and (c) both positive and negative outcomes in learning, behavior, and health. These domains provide a scaffold for the design of further research into basic developmental processes as well as a roadmap for a new era in early childhood policy focused on specific foundations of health and sources of adversity that offer promising targets for innovative intervention strategies, beginning as early as the prenatal period.

The first target area—the environment of relationships in which a young child develops—requires attention to a continuum from nurturing, responsive caregiving to neglectful or abusive interactions. This includes both family and nonfamily members as important sources of stable and growth-promoting relationships as well as critical buffers against significant threats to healthy development. The second target area—the physical, chemical, and built environments in which the child and family live—requires protection from neurotoxic exposures such as lead, mercury, and organophosphate insecticides; safeguards against injury such as the use of infant seat restraints in automobiles; and availability of safe neighborhoods and the associated social capital that supports families with young children. The third target area for intervention—appropriate versus poor nutrition—requires attention to the availability and affordability of nutritious food; parent knowledge about age-appropriate meal planning for young children that assures adequate intake of both macronutrients (e.g., protein, carbohydrates, and fat) and micronutrients (e.g., vitamins and minerals); and effective controls against the growing problem of excess caloric consumption and early obesity.

The second set of domains in the proposed framework includes a variety of physiological responses that present promising opportunities for a new generation of mediating, biological variables whose measurement might inform the effects of a broad range of experiences and interventions. In some cases, these physiological markers may be associated with specific adverse events or experiences that occur during sensitive periods in the development of the brain or other organ systems. The lifelong cognitive deficits and physical impairments associated with first trimester rubella infection or significant prenatal alcohol exposure are two prominent examples of this type of time-sensitive effect. In other circumstances, physiological changes may reflect the cumulative damage or biological "wear and tear" caused by risk factors such as recurrent abuse or chronic neglect that build up over an extended period of time. When early influences have been positive, physiological systems are typically healthy and adaptive. When influences have been adverse, systems may be dysfunctional and lead to impaired learning, maladaptive behavior, illness, disability, and a shortened life span. In both cases, genetic predisposition moderates the effects of environmental influence and differential sensitivity to context can result in differential outcomes (Obradovic et al., 2010). The identification and measurement of these physiological "footprints" offers considerable promise as a potential source of short- and medium-term markers of both resilience and vulnerability in the face of adversity, as well as potential metrics for assessing the impact of early life interventions on mediators of longer term outcomes (such as chronic diseases) that require decades to confirm. Potentially promising biomarkers that appear to be worthy of careful study include measures of oxidative stress, inflammatory cytokines, telomere length, telomerase levels, and epigenetic profiles, among many others.

The third set of target domains includes adult outcomes in educational achievement and economic productivity (high vs. low), health-related behaviors that are enhancing (e.g., nutritious diets, frequent exercise, and use of condoms to prevent sexually transmitted disease) versus those that are threatening (e.g., smoking, alcohol abuse, illicit substance use, unprotected sexual activity, antisocial behavior, and violent crime), and health status (well-being vs. disease or disorder). Although greater details regarding precise causal sequences remain to be elucidated, increasing evidence supports the conclusion that many aspects of these domains of adult competence and health can be influenced by early life experiences that are amenable to intervention.

The Evolving Context of Early Childhood Policy in the United States

The origins of public support for the notion of intervening in the lives of disadvantaged young children were fueled by four foundational influences in the 1960s—President John Kennedy's interest in mental retardation, President Lyndon Johnson's determination to eliminate the intergenerational cycle of poverty, the moral force of the civil rights movement, and emerging scholarship that challenged the prevailing view of genetic determinism and promulgated the novel idea that intelligence could be influenced by experience (Bloom, 1964; Hunt, 1961). When President Johnson called on policy makers and researchers to design a War on Poverty, they created school breakfast programs, neighborhood health centers, and a preschool summer enrichment project named Head Start, among a host of other initiatives. Over the intervening years, randomized controlled trials have demonstrated that a variety of early childhood interventions (including well-studied programs with extended longitudinal follow-up data such as the Perry Preschool Project, Abecedarian Program, and Nurse Family Partnership) can clearly make a positive difference in the development of disadvantaged young children (Campbell & Ramey, 1994; Olds, 2006; Schweinhart, 2005). After more than 40 years of concurrent advances in the science of early childhood development, the challenge facing policy makers at the end of the first decade of the 21st century is clear—it is time to leverage new scientific knowledge in the service of generating new intervention strategies that will produce substantially greater impacts.

We know more now than ever before about how young children learn and about how to facilitate the development of competencies in a variety of domains. We also have greater insights into how significant adversity can produce disruptive physiological effects on the developing brain, cardiovascular system, and immune system that can have lifelong impacts on both educational achievement and health. As the notion of early childhood policy for children and families is attracting the interest of increasing numbers of public and private sector leaders across the political spectrum, these rapidly moving scientific frontiers offer unprecedented opportunities to stimulate new thinking. With the demand for prudent investment increasing and the pressures on resource allocation growing, three challenges have become increasingly clear.

First, there is marked variability in the quality of implementation of successful programs as they are taken to scale. Demonstration projects tell us what is possible, but their replication on a broader scale has been remarkably uneven. In some cases, failure to achieve comparable impacts has been attributed to the difficulty of reproducing the charismatic leadership and exceptional motivation displayed by those who launch new programs. Others point to the inevitable shortcomings of replication efforts that attempt to serve more children at lower costs by hiring less well-trained staff and burdening them with large caseloads.

Second, increasing evidence indicates that the most effective interventions for children living in poverty produce positive outcomes, but the magnitude of their impacts is typically modest in size. This is perhaps best illustrated in 40-year follow-up data from the most frequently cited early childhood program, the Perry Preschool Project, which reported higher rates of high school graduation (from 45% to 66%) and lower rates of arrest for violent crime (from 48% to 32%) for program participants compared with a randomized control group (Schweinhart, 2005). By any reasonable standard, the magnitude of these intervention impacts would be judged as impressive. However, the fact that one third of the program participants did not graduate high school and another third were arrested indicates the need for more effective intervention strategies. Longitudinal follow-up data on recipients of home visiting services through the Nurse Family Partnership Program support a comparable conclusion (Olds, 2006). Statistically significant impacts on long-term outcomes can be achieved, but effect sizes are often modest and persistent disparities remain to be addressed.

Third, large numbers of young children and families who are at greatest risk, particularly those experiencing toxic stress associated with persistent poverty complicated by child maltreatment, maternal depression, parental substance abuse, and/or interpersonal violence, do not appear to benefit significantly from existing programs. Highly disorganized parents are less likely to seek services and more likely to drop out of programs when they do enroll. When they are successfully engaged, the needs of families facing exceedingly complex social and economic disruptions typically overwhelm conventional early childhood program staff whose expertise is restricted to child development and parenting education. Consequently, the evaluation literature on interventions for children in highly distressed families, such as programs for children who have been victims of abuse or neglect, reveals relatively limited evidence of success (MacMillan et al., 2007).

Together these three challenges underscore the need for a new era in early childhood policy and practice that is guided by science and driven by leadership that combines a strong sense of civic responsibility, an informed understanding of the positive returns that can be generated by wise investment, and a willingness to explore new ideas. Within this context, the field of early childhood intervention currently requires the concurrent pursuit of two separate, yet complementary, paths. One track leads toward closing the gap between what we know and what we do right now. Its course is well marked–enhanced staff development, increased quality improvement, appropriate measures of accountability, and expanded funding to serve more children and families. The second path heads into less well-charted territory, yet its purpose is deeply compelling–to create a new mind-set that promotes innovation, invites experimentation, and leverages the frontiers of both the biological and social sciences into transformational changes in policy and practice. The first path will bring state-of-the-art services to greater numbers of children and families. The second positions current best practices as a promising starting point, not a final destination. Both courses are essential.

Crafting a Science-Based Transition Into a New Era

In 2000, the National Research Council and Institute of Medicine produced a report entitled, ''From Neurons to Neighborhoods: The Science of Early Childhood Development'' which proposed three broad sets of research recommendations: (a) integrating child development research, neuroscience, and molecular genetics; (b) integrating the basic science of human development and the applied science of early childhood intervention; and (c) improving evaluations of early childhood interventions (Shonkoff & Phillips, 2000). Striking scientific advances over the ensuing decade now underscore the compelling need for comparable innovation in policy and practice. The following four objectives are offered as promising directions for such efforts:

The need for an expanded definition of evidence that includes well-established scientific concepts as well as benefit-cost data and the results of randomized controlled studies. The growing demand for evidence-based policies and programs is an increasingly powerful force in the early childhood policy arena. Accountability protects the interests of those who rely on services and those who provide them no less than it meets the needs of those who are responsible for funding them. The question is not whether decisions about the allocation of resources should be informed by evidence, but whether the current definition of evidence that guides early childhood investments may be too narrow. There can be no disagreement with the assertion that randomized experiments remain the gold standard for comparing the efficacy and effectiveness of alternative interventions. The value of cost effectiveness and cost–benefit assessments for calculating the monetary returns achieved from selected interventions is also not in question. That said, although these well-established sources of evidence provide useful information about existing services, they offer relatively little guidance for the compelling task of innovation.

Core concepts of development that meet the rigorous criteria of scientific peer review represent an underutilized yet equally important source of evidence for the policy arena. The extensively documented influence of early experience on gene expression is evidence that poor school performance is not an inevitable consequence of growing up in poverty. The disruptive impact of toxic stress on the development of neural circuitry in the immature brain, and the special sensitivity of the hippocampus that contains much of the neural infrastructure for simple memory and early learning, provide compelling evidence of the need for intensive intervention for very young children who have been abused or neglected. The inextricable interactions among neural circuits that subsume cognitive, language, and emotional capacities provide more than enough evidence to indicate that attention to the mental health of children experiencing significant adversity may be as important to their preparation for school success as exposure to experiences rich in language and literacy. Although evidence documenting the adverse effects of exposure to violence on brain development does not tell us anything about the effectiveness of a specific intervention program, it does underscore the need for dramatic rethinking of marginally effective programs for very young children who live in highly threatening environments. Although quantitative data from cost–benefit studies can help policy makers calculate the financial returns that can be expected from existing services, they offer relatively little guidance for creative agency directors or service providers who are motivated to develop new ideas.

The challenge facing those who seek to reduce the gap between what we know from the biological and behavioral sciences and what we do through policy and practice is to look beyond the program evaluation literature alone and leverage sound scientific concepts to drive innovation. When program evaluation data are presented favorably by advocates and dismissed as methodologically flawed by critics, responsible policy makers can still make informed decisions based on sound scientific principles, followed by ongoing evaluation to determine the impacts of their choices. In short, evidence from randomized experiments and cost–benefit studies is instructive but it simply tells us what we have learned from the past. When these data are augmented by evidence grounded in well-established scientific principles, tempered by a balance of wisdom and creativity, promising new ideas are more likely to emerge.

The need to bring a science-based approach to understanding how early childhood policies and programs can be more responsive to variations in cultural context as well as adversity associated with racial or ethnic discrimination. Acknowledgment of the importance of cultural competence in early childhood policy and practice is common, but scientific investigation of the impact of different childrearing beliefs and practices on early brain development is nonexistent. Moreover, although the need for more behavioral research into the normative development of children who grow up in a variety of cultural contexts is clear, it is essential that such research include more analyses of within-group variability rather than continue to focus primarily on between-group differences. The extent to which advances in the biology of adversity may provide greater insights into the causes of disparities in health and learning outcomes associated with minority group status also awaits further study. Recent findings that children with high neurobiological stress reactivity are more likely to exhibit maladaptive outcomes in the face of significant adversity but better adaptation in the context of low adversity suggests a particularly compelling direction for additional research in this area (Obradovic et al., 2010).

Garcia Coll et al. (1996) proposed an innovative conceptual model for developmental research on minority populations that differed from previous frameworks by placing constructs related to social position and social stratification at the core of their formulation of child development rather than at the periphery. This approach is based on the assumption that stratifications associated with racism, prejudice, discrimination, and segregation have important effects on the development of minority children, particularly with respect to the impacts of segregation, which can be manifested through various combinations of residential, economic, social, and psychological separation or frank isolation. In a related fashion, the cultural distance between professional staff in early childhood programs who are predominantly representative of the dominant culture and the increasingly diverse population of families with young children they are asked to serve presents a clear challenge that has generated considerable rhetoric, yet relatively little productive action. More empirical data on the developmental consequences of minority group status for young children, separate from the influence of social class, would promote a more constructive approach to this issue.

The need to overcome the persistent fragmentation that typifies health, education, and human services systems by leveraging an integrated, science-based framework rather than negotiating interagency agreements among conceptually disconnected programs. The initial rationale for establishing the National Research Council and Institute of Medicine Committee that produced "From Neurons to Neighborhoods" was a firm belief that the fragmented world of early childhood policy, practice, and research was guided by a single underlying science of early childhood development that needed to be articulated. As our understanding of that unified science base has deepened, the persistent disconnection that typifies the multiple policy streams that address early childhood concerns has become increasingly untenable. The contrast between research-driven attempts by departments of education to provide enriched preschool programs for low-income children and resistance by welfare departments to the establishment of high-quality standards for child care for young children whose mothers face mandated employment regulations is one striking example. Another prominent disconnection is the primary focus on physical safety and the absence of expertise in early childhood mental health within child protective services systems that have responsibility for the well-being of young children who have been abused or neglected. In these and many other circumstances, improved outcomes for children facing significant adversity are most likely to be achieved through the coordinated application of a unified, science-based framework across agencies and sectors, not through continuing attempts to foster improved interagency cooperation among disparate systems that are guided by divergent, historical traditions rather than convergent, contemporary knowledge.

The need to formulate and test new theories of change to drive more effective interventions. Early childhood policies and practices are likely to advance best within an open environment that engages a broad diversity of values and expertise, promotes intellectual flexibility and creativity, and encourages a willingness to take risks and learn from failure. This is not meant to minimize the continuing importance of efforts that focus on incremental improvements in the quality of existing programs. It is simply intended to underscore the need for dramatic rethinking in the search for more effective intervention strategies.

The biodevelopmental framework described earlier presents an integrated approach for addressing the early childhood roots of disparities in learning, behavior, and health. Advances in neuroscience suggest two currently underaddressed threats to early development that are particularly promising candidates for greater attention. The first and most compelling is the need for more effective intervention strategies to reduce the adverse biological effects of toxic stress on developing brain circuits and other organ systems. As noted above, although the underlying biology is likely to be exceedingly complex and resistant to simple remedies, there is sufficient evidence right now to support substantial new investments in innovative, relationship-based interventions for young children burdened by the stresses of child maltreatment, significant parental mental health impairments (particularly maternal depression and substance abuse), or family violence. Another candidate domain for increased attention, which is closely associated with the immediate consequences of toxic stress, is the disruptive impact of child mental health problems on early learning. In short, the provision of rich, center-based learning experiences for young children is not sufficient to prevent developmental lags if the children are burdened by anxieties or fears as a result of disruptive life circumstances that are not being addressed directly. Similarly, the provision of information on child development and advice on parenting is not sufficient for mothers and fathers with low income and limited education if the parents themselves are having considerable difficulty coping with the stresses of poverty, depression, substance abuse, food insecurity, homelessness, and/or neighborhood violence.

Stated bluntly, precipitants of toxic stress in the lives of vulnerable children and their families are not currently addressed sufficiently. Although there is a growing evidence base for effective treatment of behavioral and mental health problems in young children, the incorporation of that expertise is extremely limited in the typical early care and education center. The absence of explicit intervention strategies to address the considerable financial burdens experienced by low-income families whose children are enrolled in early childhood programs presents another serious gap, particularly in the face of evidence that the experience of living in poverty between the prenatal period and age 5 appears to have a strong association with subsequent adult earnings and work hours (Duncan, Ziol-Guest, & Kalil, 2010). Finally, the increasing racial and ethnic diversity of the early childhood population and its varied needs demands a deep commitment to the critical task of building, testing, and continually refining more sophisticated and multidimensional theories of change that speak to a broad range of childrearing beliefs and practices.

A Global Perspective on Early Childhood Development

Advances in the natural sciences generate core biological concepts that apply to all animal species. One of the central principles of this growing knowledge base is the critical role of early experiences in shaping the developmental process, whether the subjects of investigation are rodents in laboratory cages, rhesus macaque monkeys in primate research centers, or humans living in a wide variety of social, political, or economic environments. In the wealthiest countries in the world, where economic disparities, racial or ethnic discrimination, and other social challenges are associated with persistent gaps in educational achievement and health status, science can inform the development of innovative policies and practices to mitigate the negative impacts of adversity on young children. In the lowest income countries, which continue to struggle with severe malnutrition, debilitating infection, high child mortality rates, and limited educational achievement, science offers a promising opportunity to augment policies focused exclusively on child survival with a complementary array of investments in early childhood development that will help build the human capital needed to overcome deeply entrenched poverty and promote a broader social and economic agenda.

Over the past decade, the concept of early childhood development has received increasing attention in several high-profile international documents and reports, including the Education for All Dakar Declaration (UNESCO, 2000), the United Nations Millennium Development Goals (United Nations, 2000), the WHO Commission on Social Determinants of Health (Commission on Social Determinants of Health, 2008), and the 2007 *Lancet* series on child development in low-income countries, which estimated that 200 million children worldwide are not reaching their cognitive potential as a consequence of deep poverty (Grantham-McGregor et al., 2007). Concurrent with this growing global awareness, UNICEF (2009) reported that more than 30 governments have established national early childhood policies and 70 have some type of national mechanism to coordinate programs across ministries or sectors.

Despite these positive trends, however, substantial gaps remain to be closed. More than half of the world's governments do not have any policy or coordination mechanism related to early childhood health and development, and those that have articulated policies are typically guided by statements of intent rather than enforceable or implementable plans. In a parallel fashion, although the first goal articulated in the 2000 Dakar Declaration calls for "expanding and improving comprehensive early childhood care and education," it is the only goal without a quantifiable indicator or specific target by which progress can be measured. Finally, few of the major international donors have identified early childhood as a specific component in their funding strategies and most allocate less than 2% of their development assistance for education at the preprimary school level (UNESCO, 2006).

Generally speaking, global understanding of the science of early childhood development and its implications for framing policy priorities remains limited. This is further complicated by the paucity of available data to determine which interventions have measurable impacts on child well-being in a variety of political and cultural contexts. The need for coordination across policy sectors (i.e., involving ministries in health, education, and finance) as well as among the senior leadership within international agencies adds to the current challenges. Consequently, objectives like enhanced early childhood development are very difficult to achieve and sustain as they tend to be departmentalized, with child survival being the focus of the health sector and early child care and education that of the education sector. The proposed biodevelopmental model could be used to support a science-based strategy to transcend those barriers.

Concluding Thoughts

After four decades of scientific advances and early childhood program development, we have solid evidence that underscores the role of positive, early experiences in strengthening brain architecture and a growing understanding of how significant adversity damages brain circuits and undermines lifelong learning, behavior, and both physical and mental health. Moreover, neurobiology tells us that the later we wait to invest in children who are at greatest risk, the more difficult the achievement of optimal outcomes is likely to be, particularly for those who experience the early biological disruptions of toxic stress.

Complementing this knowledge base in the biological and developmental sciences, program evaluation data tell us that we can improve the life trajectories of children who face the burdens of poverty and social disadvantage, but the quality of program implementation and the magnitude of measured impacts are highly variable. This evidence base is supported further by the anecdotal reports of early childhood program staff who indicate that they see the positive impacts of their efforts on a daily basis yet are often overwhelmed by the emotional, behavioral, and social problems of many of the children and families they serve. All available information points to the same conclusion—intervention in the early years can make an important difference, and the magnitude of policy and program impacts must be increased.

Science tells us that the early childhood period is a time of both great opportunity and considerable risk, and its influence can extend over a lifetime. The foundational importance of the early years is increasingly appreciated across the political spectrum, and there is growing recognition that families, communities, the workplace, and government each has a shared interest and distinctive, noninterchangeable role to play in assuring the healthy development of all young children. Stated simply, the science of early childhood and brain development is strong and growing, the moral imperative for preventive action is compelling, and the potential social and economic returns on investment are substantial.

The distinguished biologist E. O. Wilson (1998) wrote, "We are drowning in information, while starving for wisdom. The world henceforth will be run by synthesizers, people able to put together the right information at the right time, think critically about it, and make important choices wisely.'' The challenge for informed policy making is to focus less attention on competing interpretations of program evaluation data that demonstrate statistically significant but relatively modest impacts and to direct more investment toward generating and testing new ideas about how to improve life outcomes more effectively for those whose needs are not being met. The complementary challenge for the research community is to focus less on fine-tuned measurement of what we already know about the developmental process and more on the formulation, testing, and continuous refinement of new theories of change to address significant threats in the early years of life. An exciting new era in early childhood policy, practice, and research lies at the convergence of these two agendas—an era driven by science, creativity, and pragmatic problem solving in the service of building a more humane present and more promising future for all young children and their families.

References

Barker, D. (1997). They have coronary events as adults. *New England Journal of Medicine, 335*, 1802–1809.

Barker, D., Osmond, C., Forsen, T., Kajantie, E., & Erikson, J. (2005). Trajectories of growth among children who have coronary events as adults. *New England Journal of Medicine, 353*, 1802–1809.

Bloom, B. (1964). *Stability and change in human characteristics*. New York: Wiley.

Bronfenbrenner, U. (1979). *The ecology of human development*. Cambridge, MA: Harvard University Press.

Campbell, F., & Ramey, C. (1994). Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low-income families. *Child Development, 65*(2), 684–698.

Cicchetti, D., Rogosch, F. A., Gunnar, M. R., & Toth, S. L. (2010). The differential impacts of early physical and sexual abuse and internalizing problems on daytime cortisol rhythm in school-aged children. *Child Development, 81*, 252–269.

Commission on Social Determinants of Health. (2008). *Closing the gap in a generation: health equity through action on the social determinants of health. Final report of the Commission on Social Determinants of Health.* Geneva: World Health Organization.

Davis, E. P., & Sandman, C. A. (2010). The timing of prenatal exposure to maternal cortisol and psychosocial stress is associated with human infant cognitive development. *Child Development, 81*, 131–148.

D'Onofrio, B., Singh, A. L., Iliadou, A., Lambe, M., Hultman, C., Neiderhiser, J. M., et al. (2010). A quasi-experimental study of maternal smoking during pregnancy and offspring academic achievement. *Child Development, 81*, 80–100.

Duncan, G. J., Ziol-Guest, K. M., & Kalil, A. (2010). Early childhood poverty and adult attainment, behavior, and health. *Child Development, 81*, 306–325.

Fox, S., Levitt, P., & Nelson, C. A. (2010). How the timing and quality of early experiences influence the development of brain architecture. *Child Development, 81*, 28–40.

Garcia Coll, C., Lamberty, G., Jenkins, R., McAdoo, H., Crnic, K., Wasik, B., et al. (1996). An integrative model for the study of developmental competencies in minority children. *Child Development, 67*, 1891–1914.

Garmezy, N., & Rutter, M. (1983). *Stress, coping, and development in children*. New York: McGraw-Hill.

Grantham-McGregor, S., Cheung, Y., Cueto, S., Glewwe, P., Richter, L., & Strupp, L. (2007). Developmental potential in the first 5 years for children in developing countries. *Lancet, 369*, 60–70.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science, 312*(5782), 1900–1902.

Hertzman, C. (2000). The biological embedding of early experience and its effects on health in adulthood. *Annals of the New York Academy of Sciences, 896*, 85–95.

Hunt, J. M. (1961). *Intelligence and experience*. New York: Ronald Press

Karoly, L. A., Kilburn, M. R., & Cannon, J. S. (2005). *Early childhood interventions: Proven results, future promise.* Santa Monica, CA: RAND.

Knudsen, E., Heckman, J., Cameron, J., & Shonkoff, J. (2006). Economic, neurobiological and behavioral perspectives on building America's future workforce. *Proceedings of the National Academy of Sciences of the United States of America, 103*, 10155–10162.

MacMillan, H., Jamieson, E., Wathen, C., Boyle, M., Walsh, C., Omura, J., et al. (2007). Development of a policy-relevant child maltreatment research strategy. *Milbank Quarterly, 85*, 337–374.

Meaney, M. (2010). Epigenetics and the biological definition of Gene × Environment interactions. *Child Development, 81*, 41–79.

National Scientific Council on the Developing Child. (2005). *Excessive stress disrupts the architecture of the developing brain* (Working Paper No. 3). Available at http://www.developingchild.net/reports.shtml

Obradovic, J., Bush, N. R., Stamperdahl, J., Adler, N. E., & Boyce, W. T. (2010). Biological sensitivity to context: The interactive effects of stress reactivity and adversity on socio-emotional behavior and school readiness. *Child Development, 81*, 270–289.

Olds, D. (2006). The nurse–family partnership. In N. F. Watt, C. Ayoub, R. H. Bradley, J. E. Puma, & W. A. LeBoeuf (Eds.), *The crisis in youth mental health: Early intervention programs and policies* (pp. 147–180), Westport, CT: Praeger.

O'Neal, C. R., Brotman, L. M., Huang, K., Gouley, K. K., Kamboukos, D., Calzada, E. J., et al. (2010). Understanding relations among early family environment, cortisol response, and child aggression via a prevention experiment. *Child Development, 81*, 290–305.

Pollak, S. D., Nelson, C. A., Schlaak, M. F., Roeber, B. J., Wewerka, S. S., Wiik, K. L., et al. (2010). Neurodevelopmental effects of early deprivation in post-institutionalized children. *Child Development, 81*, 224–236.

Rutter, M. (2000). Resilience reconsidered: Conceptual considerations, empirical findings, and policy implications. In J. P. Shonkoff & S. J. Meisels (Eds.), *Handbook of early childhood intervention* (2nd ed., pp. 6551–6682). New York: Cambridge University Press.

Sameroff, A. J., & Chandler, M. J. (1975). Reproductive risk and the continuum of caretaking casualty. In F. D. Horowitz, M. Hetherington, S. Scarr-Salapatek, & G. Sigel (Eds.), *Review of child development research* (Vol. 4, pp. 187–244). Chicago: University of Chicago Press.

Sameroff, A. J., & Fiese, B. H. (2000). Transactional regulation: The developmental ecology of early intervention. In J. P. Shonkoff & S. J. Meisels (Eds.), *Handbook of early childhood intervention* (2nd ed., pp. 135–150). New York: Cambridge University Press.

Schweinhart, L. J. (2005). *Lifetime effects: The High-Scope Perry Preschool Study through age 40*. Ypsilanti, MI: High/Scope Press

Shonkoff, J. P., Boyce, W. T., & McEwen, B. S. (2009). Neuroscience, molecular biology, and the childhood roots of health disparities: Building a new framework for health promotion and disease prevention. *Journal of the American Medical Association, 301*, 2252–2259.

Shonkoff, J., Phillips, D. (Eds.). (2000). *From neurons to neighborhoods: The science of early childhood development.* Washington, DC: National Academy Press.

UNESCO. (2000). *The Dakar Framework for Action: Education for all—Meeting our collective commitments.* World Education Forum, Dakar: UNESCO.

UNESCO. (2006). *Education for All Global Monitoring Report 2007. Strong foundations: Early childhood care and education.* Paris: UNESCO.

UNICEF. (2009). Policies for early childhood development. Retrieved September 10, 2009, from http://www.unicef.org/earlychildhood/index_40752.html

United Nations. (2000). United Nations millennium declaration. General Assembly Resolution 55/2

Werner, E., & Smith, R. (1982). *Vulnerable but invincible: A study of resilient children.* New York: McGraw-Hill.

Wilson, E. O. (1998). *Consilience: The unity of knowledge.* New York: Vintage.

Child Development, January/February 2010, Volume 81, Number 1, Pages 368–375

The Value of Reanalysis: TV Viewing and Attention Problems

E. Michael Foster and Stephanie Watkins
University of North Carolina

Using data from the National Longitudinal Survey of Youth (N = 1,159), this study reexamines the link between maternal reports of television viewing at ages 1 and 3 and attention problems at age 7. This work represents a reanalysis and extension of recent research suggesting young children's television viewing causes subsequent attention problems. The nonlinear specification reveals the association between television watching and attention problems exists—if at all—only at very high levels of viewing. Adding 2 covariates to the regression model eliminated even this modest effect. The earlier findings are not robust. This study also considers whether its own findings are sensitive to unobserved confounding using fixed-effects estimation. In general, it finds no meaningful relation between television viewing and attention problems.

Reanalysis is fairly common in other social sciences but relatively rare in psychology. Such analyses are often quite useful especially when the original analyses are provocative. For example, for at least three decades, demographers, economists, and sociologists have considered the impact of teenage childbearing on the lives of teen mothers and their children. Literally hundreds of studies have examined this issue, and many have focused on a handful of large, publicly available data sets, such as the Panel Study of Income Dynamics and the National Longitudinal Survey of Youth.

Our understanding of the phenomena has matured as analyses have continued. In the 1990s, provocative analyses of these data suggested that the effects had been grossly overstated and were likely small (Geronimus & Korenman, 1992); further analyses indicated that research in the 1980s and 1990s had indeed exaggerated the impact of teen childbearing but that the remaining costs were still substantial (Hoffman, Foster, & Furstenberg, 1993). Further analyses have continued to refine our understanding, emphasizing the factors that moderate the effect of early childbearing (such as a quick second birth; Kalmuss & Namerow, 1994).

This article accomplishes two purposes. First, we illustrate the importance of reanalysis with an example long of interest to developmentalists, the link between television viewing and children's behavior. A recently published article (Christakis, Christakis, Zimmerman, DiGiuseppe, & McCarty, 2004) suggests that television viewing is associated with attention deficit disorder. The first aim of this article is to reconsider whether the key findings are sensitive to modest changes in model specification. We do so by adding more covariates to the model and then by considering broader issues, such as unobserved confounding. We believe our findings demonstrate fairly convincingly that television viewing has no effect on attention problems, at least for the average child watching typical amounts of television.

The article's second contribution is to examine the effect of television viewing in a more nuanced manner. In particular, we consider whether the effect of television viewing depends on how much television a child watches using a tool underutilized in developmental research, semiparametric regression.

Semiparametric regression allows the researcher to model the effect of explanatory variables (television viewing and the potential confounding variables) on the outcome of interest, attention problems. The relation between the covariates and the outcome can be handled in one of two ways—as having a linear, parametric relation or as having a nonparametric relation. The latter involves fitting a smooth curve that can take virtually any form. That relation cannot be characterized with a single parameter estimate, but it can be plotted. These methods can be used to flexibly capture the association between variables without assuming a specific functional form and without capitalizing on chance

---

Correspondence concerning this article should be addressed to E. Michael Foster, Department of Maternal and Child Health, University of North Carolina, Chapel Hill, Rosenau Hall, Campus Box 7445, Chapel Hill, NC 27599. Electronic mail may be sent to emfoster@unc.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8101-0024

findings in the data. Such methods are increasingly common across a range of disciplines, such as biology (Huang et al., 2005), economics (Yatchew & No, 2001), health services research (Glick & Polsky, 1999), and others. We apply these methods here to assess the relation between television viewing and attention problems.

Nearly since the advent of television, researchers have been interested in its effects on children's cognitive and emotional development. In the 1960s, the National Institute of Mental Health (NIMH) commissioned $1 million worth of research projects that focused on various aspects of television and children's development (Liebert & Sprafkin, 1988). Much of that research examined the effects of television watching on children's ability to exercise self-control and maintain focus (Englander, 1997; Liebert & Sprafkin, 1988; Van Evra, 1998).

Surprisingly, very few studies have analyzed this question in large, national samples with appropriate demographic controls. A recent article by Christakis et al. (2004) spiked a resurgent interest in television's influence on children's attention and behavior. The authors assessed the link between television viewing at 1 and 3 years of age and attention problems at 7 years of age. The strengths of the study include its focus on very young children and the use of a large national sample. The authors reported a substantial—indeed, even alarming—link between early television watching and risk of attention problems at older ages. Given the levels of television viewing among children nationwide, such a strong relation would represent a major public health problem.

The Christakis et al. (2004) study raises several possibilities for reanalysis. First, the authors report only a linear analysis of the association between early television watching and later attention problems. The assumption here is that any additional hour of watching television yields the same increase in risk (i.e., in the log-odds of attention problems). Yet, research on television viewing finds that the context in which children watch television matters. Viewing behavior—duration (hours watched) and content (what is watched), with whom children watch (parents, peers, siblings), and how they watch (restricted, monitoring) are important factors to consider. The more television children watch, the more information they are exposed to and may process and retain, and the more they can replay their learned behavior (Potter, 1996; Valkenburg & Vroone, 2004; Van Evra, 1998). Bandura (2002) stated, "Heavy exposure to this symbolic world may eventually make the televised images appear to be the authentic state of human affairs" (p. 137).

Second, Christakis et al. (2004) included an array of covariates in their regressions, attempting to rule out spurious, third variable problems. Yet, they failed to include two available measures that we think are essential and are readily available in the data set. These two variables include a measure of mothers' achievement, assessed with the Armed Forces Qualification Test (AFQT), and a measure of the family's income. Mothers with lower levels of achievement or income may allow their children to watch more television, and these characteristics may influence behavior in later childhood. Whether the results Christakis et al. reported are robust to the additional inclusion of these two important variables represents a prime opportunity for reanalysis.

Method

Data

The data for these analyses and the original study were drawn from a large, nationally representative sample of children and their families followed longitudinally, the 1979 cohort of the National Longitudinal Survey of Youth (NLSY79). Participants in the original study were ages 14–24 in 1979, and as these women reached childbearing age, the study added an extensive battery of assessments involving their own children. Numerous analyses have used these data to consider a range of important questions involving children's development.

Of participants in the current study, 82% were non-Hispanic White, 6% were Hispanic, and 12% were African American when weighted (unweighted percentages were 62%, 14%, and 24%, respectively). In terms of socioeconomic status, 44% of participants reporting having a high school degree or less. (This figure was calculated with the sampling weights. The unweighted figure is 48%.) The sample is divided equally by gender.

The Christakis et al. (2004) study examined the link between television viewing between ages 1 and 3 and attention problems at age 7. Attention problems are assessed with the five-item hyperactivity subscale of the Behavior Problems Index, a maternal-report measure of children's behavior available in the data. We followed these authors in creating a binary classification representing attentional "problems" as either present or absent, using a cut-point of 120 on the same-gender standardized hyperactivity score. This procedure yields a prevalence for attentional problems that is similar to published reports of attention deficit hyperactivity disorder prevalence among young children in community samples (Wolraich, Hannah, Pinnock, Baumgaertel, & Brown, 1996).

Overview of Semiparametric Regression

Semiparametric regression is a special case of nonparametric regression. Nonparametric methods allow the researcher to examine the association between two variables without specifying the form of the relation a priori. Unlike parametric regression, which describes the relation of the data using a regression equation, nonparametric regression captures the relation using a smooth curve that can take virtually any shape. This relation can be described graphically. In essence, the curve represents a weighted, moving average of values of the dependent variable across the range of the explanatory variable.

Even though we are letting the data "speak for themselves," one must be cognizant of not underfitting or overfitting the association. A span parameter, the percentage of data points used to calculate the moving average, is set for the model to address this issue. Past experience suggests that if the parameter is too low (0.15–0.35), the graph may display excessive noise. If the parameter is set too high (0.75–1.00), important fluctuations in the data may be missed. An intermediate span of 0.65 is suggested as a compromise between these two extremes (Jacoby, 2000).

In our application, we used a semiparametric version of this model that allowed us to include key covariates in a nonparametric manner and then to control for other covariates as linear predictors. In our case, the group of covariates entered linearly were those we were interested in not for their own sake but as potential confounders (such as family income; Fox, 2000) We included these variables parametrically in order to provide the readers with a parametric estimate of the link between that covariate and the outcome.

The semiparametric model is not only a descriptive tool but also can be used to draw statistical inference. Partial F tests are available to test for nonlinearity of selected covariates as well as interaction between covariates. By contrasting two models where one covariate enters linearly and the other enters nonparametrically, one can derive an F statistic to test for the significance of a partial nonlinear relation. This principle is extended to test for the significance of interaction terms (Fox, 2000). (Note that we might have included key covariates [like family income] nonparametrically as well. Doing could be expected to further diminish the predictive power of television viewing. If the residual relation between television viewing and attention problems had remained significant, we would have explored these models. As illustrated below, simply including the covariates linearly was enough to make the putative effect of television viewing disappear.)

Results

Replicating the Results in Christakis et al. (2004)

We start by following the methods described in Christakis et al. (2004) to replicate the authors' findings (and correcting for minor problems in the original analysis). We obtain sample sizes, means, and standard deviations for the key variables that are nearly identical to those in Christakis et al., and when dissimilar, differ only in the first decimal place.

Our "age 1" sample was 1,288 children (compared to 1,278 as published in the Christakis et al., 2004, report), while our "age 3" sample was 1,404 (compared to 1,345). One difference in our methods is that we used the sampling weights for the year 2000 for all of the children in the study instead of using the sampling weights from the year the outcome was measured as in the original study. (While the effect on the parameter estimates was negligible, using the weights from a single wave of data collection produces results that are more clearly representative of the population sampled.) Doing so decreased the sample size for the "age 1" regressions to 1,039 children and the "age 3" regressions to 1,159.

We next replicated the multivariate logistic regressions presented in the original article. The odds ratios for the continuous measure of television viewing in the regressions are shown in Models 1 and 3 of Table 1. At both ages 1 and 3, the effect of an additional hour of daily television on the probability of attention problems is very small—one half of 1%.

These largely replicate the results presented in the original report (Christakis et al., 2004, Table 2). These findings suggest that our sensitivity analyses began at the right starting point—the estimates in the original study.

Table 1
Odds Ratios for Various Regression Analyses

		Age 1				Age 3			
		Original		Extended		Original		Extended	
Age of child covariates model		1	2	3	4	5	6	7	8
Specification of television effect									
A. Continuous viewing	OR	$1.08^{\dagger}$	—	1.06	—	1.09*	—	1.05	—
	SE	0.04		0.05		0.04		0.05	
	ME	$0.005^{\dagger}$		0.004		0.005*		0.002	
B. Viewing categories									
0 hr (omitted)	OR	—	REF	—	REF	NA	NA	NA	NA
	SE								
	ME								
Between 0 and 1 hr	OR	—	0.94	—	0.80	—	REF	—	REF
	SE		0.53		0.50				
	ME		−0.004		−0.013				
Between 1 and 2 hr	OR	—	1.25	—	1.25	—	0.44	—	0.39
	SE		0.47		0.52		0.26		0.25
	ME		0.02		0.01		$-0.04^{\dagger}$		$-0.04^{\dagger}$
Between 2 and 3 hr	OR	—	1.04	—	1.04	—	0.87	—	0.77
	SE		0.42		0.44		0.45		0.43
	ME		0.002		0.002		−0.008		−0.01
Between 3 and 4 hr	OR	—	1.89	—	1.82	—	1.17	—	0.82
	SE		0.80		0.82		0.61		0.46
	ME		0.05		0.05		0.009		−0.01
Between 4 and 5 hr	OR	—	1.75	—	1.99	—	0.64	—	0.49
	SE		0.88		1.03		0.35		0.30
	ME		0.05		0.06		−0.02		−0.03
Between 5 and 6 hr	OR	—	0.62	—	0.74	—	1.21	—	0.83
	SE		0.49		0.60		0.74		0.57
	ME		−0.03		−0.02		0.01		−0.01
Between 6 and 7 hr	OR	—	1.32	—	1.24	—	4.01*	—	3.29
	SE		1.05		1.18		2.45		2.53
	ME		0.02		0.015		0.14		0.10
Greater than 7 hr	OR	—	2.04	—	1.58	—	1.94	—	1.41
	SE		0.94		0.90		1.10		0.92
	ME		0.06		0.03		0.05		0.02
n		1,039	1,033	892	887	1,159	1,143	946	934

Note. These analyses use 2000 sampling weights for the children. Odds ratios (OR) are presented with robust standard errors (*SE*) below in parentheses. Also reported are the marginal effects (ME) of the covariate on the predicted probability of attention problems. For the models labeled "original covariates," covariates include child gender and age, race, gestational age of child, alcohol and tobacco consumption during pregnancy, number of children in household, presence of two parents, cognitive stimulation and emotional support scores, mother's age, urbanicity, year fixed effects, maternal education, maternal depression, and maternal self-esteem. The "extended covariates" include mother's Armed Forces Qualification Test score and family poverty status.
$^{\dagger}p < .10$. $^{*}p < .05$.

Extension: Adding More Covariates

Our first concern with the initial analyses involved the role of various family factors that might be confounded with the effect of television watching. The NLSY79 data are particularly well suited to these analyses because they provide extensive information on the children's development and their environment. These analyses involved a series of regressions that controlled for additional background characteristics of the mother: her academic achievement and a measure of family income. Mothers' academic achievement was measured with the AFQT percentile score. The AFQT, which was administered to NLSY sample members in 1980, assesses paragraph comprehension, arithmetic reasoning, word knowledge, and mathematics knowledge. The AFQT is described by

Neal and Johnson (1996) as a measure of basic skills, or human capital, attained. While the original analysis controls for maternal education, the AFQT is an important measure of maternal skills and may be an important influence on how much television the mother allows the child to watch.

Second, we include a measure of the poverty status of the family at ages 1 and 3 in the respective regressions. Poverty status captures a family's ability to purchase the resources and goods (e.g., schools, housing, food, safe, and cognitively enriched learning environments) that are critical for successful development. It may also be an important influence on how much television the child is allowed to watch.

In the age 1 regressions, once these two characteristics of the mother and the family were included, the continuous measure of television viewing (i.e., as in the original analysis) was no longer statistically significant (compare Model 1 vs. Model 2). Models 3 and 4 report similar analyses for television viewing at age 3.

Extension: Fixed-Effects Estimation and the Potential for Unobserved Confounding

The analyses presented above consider the effect of television viewing conditional on observed family factors. As a reviewer noted, it seems quite possible (if not likely) that our list of covariates is incomplete and that other third factors remain—these might involve characteristics of the child, his or her family, and/or his or her neighborhood that affect television viewing and attention problems. These variables likely are positively correlated with both television viewing and attention problems—i.e., if anything we suspect the estimate we have reported above is an overestimate.

However, because these other confounders are unmeasured or omitted, whether and how they influence the observed relation between television viewing and attention problems is unknown. Some unobserved factor, however, may encourage television viewing yet reduce attention problems. In that case, the effect estimated here may underestimate the effect of television viewing.

While unobserved factors cannot be measured, they can be accounted for in estimation. One way to do so involves fixed-effects estimation (Greene, 2008). That method has been used in a range of developmental analyses (e.g., Foster & Kalil, 2007). In essence, this method allows each child to have his or her own intercept. As a result, the relation between variables is based on *within-child*, intertemporal variation only. As a result, this method effectively eliminates cross-child variation in unobserved, time-invariant factors that may confound the relation of interest.

We conducted such analyses using data from ages 1, 3, 5, and 7. We examined the effect of lagged television viewing on attention problems. For these analyses, we used an extension of the model structure employed above—in particular, we used the twice-lagged measure of television viewing (from ages 1 and 3) to explain the outcome two waves later (ages 5 and 7). In essence, this method involves assessing the effect of the change in early television viewing between ages 1 and 3 on the change in attention problems between ages 5 and 7. The analysis also included changes in other time-varying characteristics as regressor (such as the change in poverty status). Estimation was performed using fixed-effects logistic regression. (We also conducted the analyses with once-lagged television viewing, and the results were unaffected.)

In the fixed-effects model, the point estimate of the effect of television actually was negative (–.04). The effect implies that television viewing may reduce attention problems but was not statistically significant (p = .32). (The implied marginal effect was roughly comparable to that in the earlier analyses.)

To some extent, the lack of significance is driven by a lack of power—statistical power in the fixed-effects regression is driven by individuals who change values on the explanatory and dependent variables over time. For a fairly stable construct, like attention problems, power is limited. (The reported standards are not incorrect, just larger.) Power was sufficient enough, however, to reject the null hypothesis that the random- and fixed-effects estimates were equivalent. (This test is known as the Hausman test.) Rejection in this case suggests that the random-effects estimates are indeed biased by unobservables—in this case, suggesting that other child and family disadvantages are correlated with television viewing and inflate its effect in standard analyses (like those in the original Christakis et al., 2004, article).

Note that because these analyses relied on all waves of data, we analyzed multiply imputated data to preserve cases. In doing so, we assume the data are missing at random with regard to both the measured covariates and the time-invariant unobservables (Foster, Fang, & the Conduct Problems Prevention Research Group, 2004). We used Rubin's rules to combine the estimates from three imputations. (Little & Rubin, 1987, 2002). In imputing the data we preserved the hierarchical structure of the data (i.e., the nesting of time points within individuals) by imputing a "wide" file—that is, by treating each wave's variables as a separate variable on a single, individual-level record.

Extension: Nonlinear Effects of Television Viewing

Our analysis described the bivariate relation between early television exposure and attention problems with a semiparametric additive model. This model allowed us to enter television viewing without an underlying functional form while still having the flexibility to include linear covariates. The smoothing parameter was set at 0.65.

The semiparametric model revealed a minimal linear increase in attention problems as average daily television viewing increased from 5 to 15 hr while controlling for all other linear covariates in the model (see Figure 1). However, the estimate was less precise for the upper levels of television viewing. The majority of the data, represented by tick marks on the horizontal axis, were located between approximately 0 and 8 hr of television viewing. Thus, the 95% confidence band, which is a measure reflecting the precision of the behavioral score estimate, was quite large for children exposed to greater than 5 hr of television.

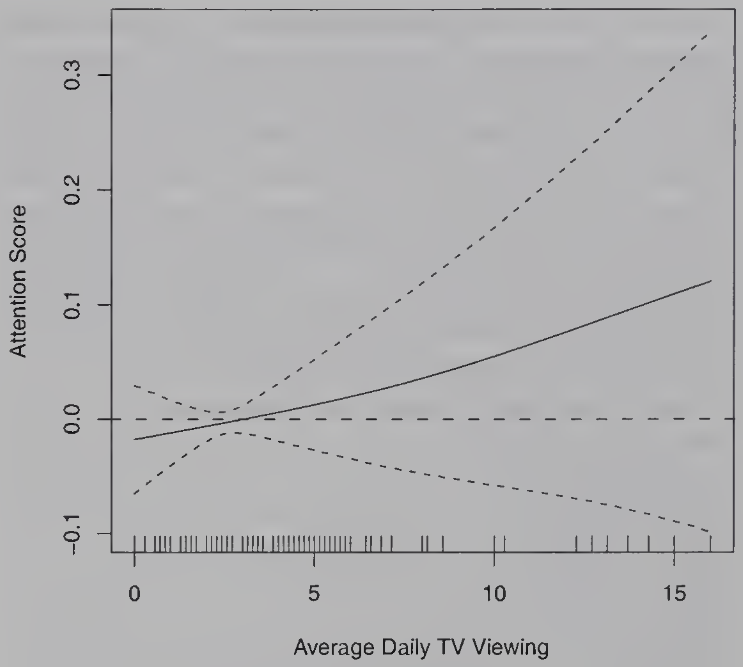


Figure 1. Semiparametric analysis of the association between average daily television viewing and attention problems. These data taken from the 1979 cohort of the National Longitudinal Survey of Youth represent a smooth function summarizing the dependence of attention score on average daily television while controlling for covariates linearly. Covariates for alcohol and tobacco were combined to fit the model on the logit scale. No differences in the plots were observed when these covariates were fit independently or combined using the normal scale.

The results of the semiparametric estimation allowed us to test selected hypotheses. First, when comparing the fully linear and semiparametric models, the test of nonlinearity was not statistically significant (p = .10). That is, we could not reject the null hypothesis that television viewing has a linear effect. However, we also were unable to reject the null hypothesis that television has no effect at all (p = .10). One also can see that even at high levels of television viewing the confidence band includes zero (no effect). As one would expect the size of the band broadens at higher levels of viewing because of the limited number of observations at those levels.

Table 2 presents the full set of estimates for this model. The estimates include the parametric estimates for those covariates included parametrically. These estimates were used to adjust the outcome for the effect of these covariates.

Table 2
Semiparametric Model: Parameter Estimates

Variable	Parameter estimate	Standard error	*p* value
Boy	0.56	0.45	.22
Black	−1/50	0.96	.12
Hispanic	0.50	0.63	.43
Child's age	0.01	0.07	.92
Mom's age	−0.04	0.12	.74
Mom's education	−0.14	0.12	.23
Children in the household	−0.55	0.27	.04*
Two-parent household	0.55	0.80	.49
Gestational age	−0.05	0.13	.73
Self-esteem	−0.06	0.06	.28
Depression	0.02	0.03	.49
Alcohol, tobacco	0.42	0.32	.19
Urbanicity (omitted category)			
Urbancity A	−0.19	0.79	.81
Urbancity B	−1.42	0.74	.05
Urbancity C	−2.01	1.14	.08
Year (omitted category)			
1996	1.43	2.27	.53
1998	2.13	2.27	.35
2000	2.10	2.46	.39

Note. All estimates rounded to two decimal places.
$^*p < .05$.

Discussion

More young children, that is, infants, toddlers, and preschoolers, watch television today. American children in particular watch a lot of television: on average 2.2 hr per day at age 1 and 3.6 hr by age 3 (Christakis et al., 2004). The American Academy of

Pediatrics has recently recommended that children under 2 watch no television and that older children's television watching be limited (American Academy of Pediatrics, 2001). There are many good reasons to think that this idea is a good one. For example, the time that children spend watching television might more productively be spent in less sedentary activities. Our work, however, questions the claim that television viewing causes a serious mental disorder, attention deficit disorder.

Our reanalyses produced two main findings. First, the association between early television watching (at ages 1 and 3) and risk of attention problems measured at age 7 is significant—if at all—only for the small group of children (10%) of the sample who watch 7 or more hr of TV per day. In other words, the relation is nonlinear, and it is not true, as Christakis et al. (2004) were quoted in the print media as saying, that each extra hour of television viewed per day is associated with a 10% increased likelihood of developing attention problems later on. Modest levels of television viewing do not appear to be detrimental, even for young children.

Second, and even more importantly, our analyses reveal that even this relation becomes nonsignificant once comparisons of children watching differing amounts of television are adjusted for two additional factors. In particular, we estimate regressions that control for maternal academic achievement and the child's poverty status in early life.

Both of these characteristics increase the likelihood that children watch more television and also influence child outcomes. As a result, analyses that ignore these relation confound the effects of television watching with that of these other factors. For example, mothers who are less oriented toward academic achievement might be more likely to allow their young children to watch TV in lieu of more enriching activities. Poor children might have fewer opportunities to participate in activities outside the home, and they, too, might be more likely to watch TV. As a result, the effects of poverty are mistakenly attributed to television watching.

We recognize that both the original study and this study may neglect the role of nonlinearities. For these analyses, we rely on semiparametric regression to capture the association between early television exposure and attention problems. These analyses provide little evidence of a link between attention problems and television viewing, linear or otherwise. We consider whether and how the amount of television viewing matters using semiparametric methods. Those analyses suggest that the linear model fits the data adequately—there was no relation between television viewing and attention problems. The lack of a relation between television viewing and attention problems in our linear model did not mask a nonlinear relation.

Christakis et al. (2004) conclude that "the magnitude of the risk associated with television viewing . . . is clinically significant." The analyses presented here suggest that the relation between television viewing and behavior is complex, *but the weight of the evidence suggests that reports of a strong relation were premature.* If there are effects of television viewing, they occur principally at high levels with the probability of effect varying between age groups. The average 2- to 17-year-old watches 2.8 hr of television per day, and the average American watches just over 4 hr per day Center for Screentime Awareness, n.d.). We find negative effects of television viewing occur only at much higher levels.

It is worth noting that there are potential statistical problems that may mask the negative effects of television viewing. For example, the measure of television viewing was reported by mothers, and those reports may not be especially accurate. In that case, measurement error would work to attenuate estimates of the parameter of interest. However, social desirability may lead parents to underreport the amount of television their children watch. If more advantaged parents are more embarrassed by how much television their children watch, then maternal reports would be negatively correlated with parental advantage. In that case conventional analyses like those in the original study would overestimate the effect of television.

Data problems may further mask the effect of television viewing. As discussed above, the effect of television may depend on the "quality" of viewing—with whom, the content of the shows, and so on. Some research suggests these dimensions of viewing may matter as much as quantity. When Wright examined television content and educational skills, children who watched educational television actually had *higher* school readiness and language skills. Yet children who frequently viewed general programming performed more poorly on these measures (Wright et al., 2001). Christakis et al. (2004) also assessed the effect of television content on children's behavior; exposure to violent content posed a significant risk of antisocial behavior later in childhood. Of course, these studies are only suggestive at this point. They suffer from the same methodological limitations as research into the quantity of television viewing, and the selection issues are at least as great.

A standard assessment of omitted variable bias suggests that our estimate of the effect of quantity is the effect of quantity per se plus the effect of quality times the association between quality and quantity (Greene, 1993). In other words, if children who watch more television watch *worse* television, then the effect estimated here captures the effect of both quality and quantity. In that sense, our estimate may exaggerate the effect of the quantity of television per se (i.e., is greater than the effect of quantity holding quality constant). That effect is small, but television may remain detrimental for children who watch a lot of certain types of television.

Clearly, our understanding of the effects of television viewing is still evolving, and so the broader lessons from this article do not involve the effect of television viewing per se. Our intent is to highlight the importance of reanalysis and data sharing so that researchers may strengthen the knowledge base for social science more generally. When research findings that have the potential to influence public health policy are translated into guidelines for parents, and/or affect parental beliefs, reanalysis is vital to ensure that the findings are robust. Transparency of data methods and results ensures confidence in study findings and prevent misleading media coverage.

References

American Academy of Pediatrics. (2001). American Academy of Pediatrics: Children, adolescents, and television. *Pediatrics, 107,* 423–426.

Bandura, A. (2002). Social cognitive theory of mass communication. In J. Bryant & D. Zillmann (Eds.), *Media effects: Advances in theory and research* (2nd ed., pp. 121–153). Mahwah, NJ: Erlbaum.

Center for Screentime Awareness (n.d.). *TV turnoff network.* Retrieved January 1, 2009, from http://www.screentime.org/index.php?option=com_content&task=view&id=7&Itemid=10

Christakis, D. A., Zimmerman, F. J., DiGiuseppe, D. L., & McCarty, C. A. (2004). Early television exposure and subsequent attentional problems in children. *Pediatrics, 113,* 708–713.

Englander, E. K. (1997). *Understanding violence.* Mahwah, NJ: Erlbaum.

Foster, E. M., Fang, G. Y., & the Conduct Problems Prevention Research Group. (2004). Alternative methods for handling attrition in prevention research: An illustration using data from Fast Track. *Evaluation Review, 28,* 434–464.

Foster, E. M., & Kalil, A. (2007). Living arrangements and children's development in low-income White, Black, and Latino Families. *Child Development, 78,* 1657–1674.

Fox, J. (2000). *Multiple and generalized nonparametric regression.* Thousand Oaks, CA: Sage.

Geronimus, A. T., & Korenman, S. (1992). The socioeconomic consequences of teen childbearing reconsidered. *Quarterly Journal of Economics, 107,* 1187–1214.

Glick, H. A., & Polsky, D. (1999). Analytic approaches for the evaluation of costs. *Hepatology, 29*(Suppl. 6), 18S–22S.

Greene, W. H. (1993). *Econometric analysis* (2nd ed.). New York: Macmillan.

Greene, W. H. (2008). *Econometric analysis* (6th ed.). Upper Saddle River, NJ: Prentice Hall.

Hoffman, S. D., Foster, E. M., & Furstenberg, F. F., Jr. (1993). Re-evaluating the costs of teenage childbearing. *Demography, 30,* 1–13.

Huang, L. S., Cox, C., Myers, G. J., Davidson, P. W., Cernichiari, E., Shamlaye, C. F., et al. (2005). Exploring nonlinear association between prenatal methylmercury exposure from fish consumption and child development: Evaluation of the Seychelles Child Development Study nine-year data using semiparametric additive models. *Environmental Research, 97,* 100–108.

Jacoby, W. G. (2000). Loess: A nonparametric, graphical tool for depicting relationships between variables. *Electoral Studies, 19,* 577–613.

Kalmuss, D. S., & Namerow, P. B. (1994). Subsequent childbearing among teenage mothers: The determinants of a closely spaced second birth. *Family Planning Perspectives, 26,* 149–159.

Liebert, R. M., & Sprafkin, J. (1988). *The early window: Effects of television on children and youth* (3rd ed.). Elmsford, NY: Pergamon Press.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data.* Hoboken, NJ: Wiley.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). Hoboken, NJ: Wiley.

Neal, D. A., & Johnson, W. R. (1996). The role of premarket factors in Black–White wage differences. *Journal of Political Economy, 104,* 869.

Potter, W. J. (1996). Considering policies to protect children from TV violence. *Journal of Communication, 46,* 116–138.

Valkenburg, P. M., & Vroone, M. (2004). Developmental changes in infants' and toddlers' attention to television entertainment. *Communication Research, 31,* 288–311.

Van Evra, J. (1998). *Television and child development* (2nd ed.). Mahwah, NJ: Erlbaum.

Wolraich, M. L., Hannah, J. N., Pinnock, T. Y., Baumgaertel, A., & Brown, J. (1996). Comparison of diagnostic criteria for attention-deficit hyperactivity disorder in a county-wide sample. *Journal of the American Academy of Child and Adolescent Psychiatry, 35,* 319–324.

Wright, J. C., Huston, A. C., Murphy, K. C., St Peters, M., Pinon, M., Scantlin, R., et al. (2001). The relations of early television viewing to school readiness and vocabulary of children from low-income families: The early window project. *Child Development, 72,* 1347–1366.

Yatchew, A., & No, J. A. (2001). Household gasoline demand in Canada. *Econometrica, 69,* 1697–1709.

Child Development, January/February 2010, Volume 81, Number 1, Pages 376–389

Counterfactual Reasoning: Developing a Sense of ''Nearest Possible World''

Eva Rafetseder, Renate Cristi-Vargas, and Josef Perner
University of Salzburg, Austria

This study investigated at what point in development 3- to 6-year-old children begin to demonstrate counterfactual reasoning by controlling for fortuitously correct answers that result from basic conditional reasoning. *Basic conditional reasoning* occurs when one applies typical regularities (such as ''If 'whenever' it doesn't rain the street is dry'') to counterfactual questions (such as ''If it had not rained, would the street be wet or dry?'') without regard to actual events (e.g., if street cleaners had just been washing the street). In *counterfactual reasoning*, however, the conditional reasoning must be constrained by actual events (according to the ''nearest possible world''). In situations when counterfactual reasoning and basic conditional reasoning would yield the same answers, even the youngest children gave mostly correct answers. However, tasks in which the 2 reasoning strategies resulted in different answers proved unusually difficult even for the older children.

People often reflect on how an event in the past might have turned out differently. Creating such alternatives to known facts is called counterfactual reasoning (Roese, 1997). Although it is common among adults, it is still not clear what is necessary for proper counterfactual reasoning. This is a serious shortcoming because counterfactual reasoning is relevant in many different areas such as cognitive psychology, social psychology, developmental psychology, and philosophy.

Developmental psychologists have addressed this issue by exploring whether children's counterfactual reasoning abilities are related to their understanding of causation (German, 1999; Harris, German, & Mills, 1996; Kavanaugh & Harris, 2000), their understanding of false belief (Grant, Riggs, & Boucher, 2004; Guajardo & Turley-Ames, 2004; Müller, Miller, Michalczyk, & Karapinka, 2007; Perner, Sprung, & Steinkogler, 2004; Riggs, Peterson, Robinson, & Mitchell, 1998), their feeling of regret and relief (Amsel & Smalley, 2000; Amsel et al., 2003; Guttentag & Ferrell, 2004), their understanding of counterfactual and actual worlds as alternative possibilities at a certain time in the past (Beck, Robinson, Carroll, & Apperly, 2006; Byrne, 2005), or their executive functions such as inhibitory control (Beck, Riggs, & Gorniak, 2009) and working memory (German & Nichols, 2003; Robinson & Beck, 2000).

The majority of these developmental studies assessed the ability to reason counterfactually by telling a short story (e.g., ''Carol made the floor all dirty with her shoes'') and then asking a subjunctive question about the past (or present): What if Carol had taken her shoes off—would the floor be dirty? (Harris et al., 1996). Children's main error was to answer with what was actually the case in the story, that is, that the floor was dirty (referred to as a ''reality error''). The predilection for reality errors subsides between 3 and 5 years, although the exact age varies across studies. However, in studies that assessed the reasoning differently (such as showing counterfactual emotions like regret), children showed difficulties with some counterfactual tasks until 6 years or older (Amsel et al., 2003; Amsel & Smalley, 2000; Beck et al., 2006; open counterfactuals; Guttentag & Ferrell, 2004). This discrepancy reveals a shortcoming in existing studies: They fail to distinguish between the kind of reasoning the experimenter intended to investigate and the kind of reasoning the children employed.

Answering a subjunctive past question correctly shows that children can reason with conditionals (e.g., if X is F then Y is G) and that they can do so independently of what the actual states of X and Y

We would like to express our appreciation to the heads and children of the following kindergartens for their cooperation and valuable time participating in this project: Betriebskindergarten St. Johanns-Spital, Pfarrcaritaskindergarten Gallneukirchen, Vereinskindergarten Perg Haydnstraße, and Vereinskindergarten Perg Friedhofstraße. This research was financially supported by Austrian Science Fund project (I140-G15) ''Counterfactual Reasoning in Children'' as part of the ESF EUROCORES LogiCCC initiative and forms part of the doctoral dissertation of Eva Rafetseder in the Department of Psychology at the University of Salzburg.

Correspondence concerning this article should be addressed to Eva Rafetseder, Department of Psychology, University of Salzburg, Hellbrunnerstraße 34, A-5020 Salzburg, Austria. Electronic mail may be sent to eva.rafetseder@sbg.ac.at.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8101-0025

are. This demonstrates an ability to entertain counterfactual states, but that ability is not the same as counterfactual reasoning. All *kinds* of conditional reasoning can depend on assumptions that are not currently true; this is not limited to counterfactual reasoning. For instance, the developmental literature considers a question about the future—"If X will be F then Y will be . . .?"—to demand future hypothetical reasoning (Riggs et al., 1998), and studies have consistently found these questions to be much easier than subjunctive questions about the past. Another kind of conditional, rarely discussed in the developmental literature, is the universally quantified conditional expressing a general regularity, which is typically formulated in the present tense indicative. "If X is F then Y is G" can be read to mean that whenever X is F then Y is G, without reference to any particular instance at a particular time. Reasoning with such universals demands what we call "basic conditional reasoning."

The problem with previous investigations is that although they used a subjunctive past question, children might have applied basic conditional reasoning and arrived at the same answers. For instance, when an investigator asks out of the blue, "If Carol had taken off her shoes, would the floor be dirty or clean?" a child could answer with "(probably) clean" by use of basic conditional reasoning, applying "If 'whenever' people take off their shoes floors tend to stay clean" to the case of Carol. Children often interpret experimenters' instructions not as intended but in a pragmatic way. For instance, Rumain, Connell, and Brain (1983) pointed out that children interpret "if" pragmatically. They often interpret the statement "If you mow the lawn, then I will give you five dollars" to mean "If you do not mow the lawn, then I won't give you five dollars." In other words, they try to figure out what the speaker actually means by the sentence. On the basis of these assumptions, they draw conclusions, for example, that if they do not mow the lawn, they will not earn money. Although this conclusion is not logically valid, it would be true in most situations. Basic conditional reasoning is, of course, only a minimal instance of reasoning because it consists of two simple processes: (a) memory retrieval, which is at best a prerequisite for logical reasoning (Markovits & Barrouillet, 2002), not an act of reasoning in itself, and (b) the simple logical reasoning rule of universal instantiation.

What abilities do children need to demonstrate proper counterfactual reasoning? Lewis (1973) proposed that, unlike normal conditionals whose truth depends on the real world, the truth of counterfactual claims is determined by what is true in the *nearest possible world* (that which resembles the real world most closely). Consider the sentence "If Carol had taken her shoes off, the floor would have stayed clean." This is true only if, in a possible world in which Carol had taken her shoes off (but is otherwise as similar as logically possible to the actual world), the floor would have stayed clean. The counterfactual claim is false if there is some nearer possible world, where Carol removed her shoes but the floor was still dirty—if, for instance, before Carol entered the room with her dirty shoes, her nasty brother had already messed up the floor with *his* dirty shoes. Hence, in this nearest possible world, even though Carol takes off her shoes, the floor is still dirty. With counterfactual reasoning, we would thus conclude that the floor would have been dirty, a conclusion that basic conditional reasoning would not have produced. We think that developmental investigations must distinguish more clearly between counterfactual and basic conditional reasoning.

There is consistent evidence in the developmental literature that children become capable of answering indicative future questions (requiring future hypothetical reasoning) before 3 years of age. This is reliably earlier than they can answer past subjunctive questions (Beck et al., 2006; Perner et al., 2004; Riggs & Peterson, 2000; Riggs et al., 1998; Robinson & Beck, 2000; Study 1), which tend to elicit the reality error. There are several points of difference that may make counterfactual reasoning particularly difficult, especially for children. First, the counterfactual assumption, by definition, contradicts with the corresponding fact (Carol took her shoes off—Carol was wearing her shoes), whereas in future hypothetical reasoning the corresponding fact in the future is not yet known. Second, the use of the subjunctive mood signals that the real sequence of events cannot be ignored, but must play a role in the counterfactual reasoning process. This requires that two different models of the world stay simultaneously active (Beck et al., 2006; Byrne, 2005), which requires more information-processing capacity and is likely to intensify the interference of known facts with counterfactual assumptions more than if hypothetical assumptions happen to differ from reality. Researchers have amply documented the problems created by the interference of reality with the process of counterfactual assumption and conclusion for very young children of 2½–5 years (Beck et al., 2006; Perner et al., 2004; Riggs & Peterson, 2000; Riggs et al., 1998; Robinson & Beck, 2000, Study 1). Finally, the real sequence of events that is being counterfactually altered must be kept active for a good reason. It plays an important role in how the counterfactual model needs to be constructed, which developmental psychology has yet to investigate but is hotly debated in philosophy.

The last point may especially explain the existing discrepancy in study results about the onset of counterfactual reasoning. Children master some counterfactual problems (past subjunctive questions) very early, at the age of 3½–4½ years (Beck et al., 2006; standard counterfactuals, German, 1999; German & Nichols, 2003; Harris et al., 1996; Perner et al., 2004; Riggs et al., 1998), but some investigations show little sign of counterfactual thinking before the age of 6 years (Amsel et al., 2003; Amsel & Smalley, 2000; Beck et al., 2006; open counterfactuals, Guttentag & Ferrell, 2004).

One way of making sense of these discrepancies is to assume that even the youngest children are able to reason counterfactually but are prevented from showing their prowess on more difficult problems; scholars have suggested two possible reasons for this. One is that children (German, 1999; Guttentag & Ferrell, 2004), like adults (Roese, 1997), tend to be more likely to engage in counterfactual reasoning for negative outcomes than for positive outcomes. However, Kavanaugh and Harris (2000; Harris, 2000) tested this hypothesis directly and found no difference in children's ability to use counterfactual reasoning for negative and positive outcomes. The other suggestion comes from German and Nichols (2003), who provided evidence that because of children's working memory limitations, the length of the inferential or causal chain from the counterfactual assumption to the predicted outcome affects children's ability to engage in counterfactual reasoning.

In contrast, Perner and Rafetseder (in press; Perner, 2000) worked from the premise that younger children *cannot* reason counterfactually and suggested that these children's correct answers to past subjunctive questions might be based on basic conditional reasoning (using our terminology). Depending on the particular task, children were more or less likely to make suitable background assumptions leading to the same answer that counterfactual reasoning would produce, which made their answers look like evidence for counterfactual reasoning. The early use of basic conditional reasoning may explain why some studies show little sign of counterfactual reasoning before the age of 6 years, provided that we can show that counterfactual reasoning takes the nearest possible world constraint into account. To provide more decisive evidence for counterfactual reasoning in children, we designed stories in which counterfactual reasoning produces different answers than basic conditional reasoning does.

Experiment 1

The principal objective of this study was to determine whether children can answer "counterfactual" subjunctive questions about a past event correctly when a basic conditional reasoning approach yields a different answer. We used two different story worlds, the *sweets story* and the *dwarf story*. In each story world, four different sequences of events could take place, and each sequence comprised two component events, in this case, two transformations of an object's location. For instance, in the sweets story, mother puts sweets regularly on either the top shelf or the bottom shelf (the first transformation), and then either the tall boy or the little girl comes looking for the sweets and then takes them into his or her room (the second transformation). The tall boy can reach both shelves, so he can take the sweets into his room regardless of where mother puts them. The little girl, however, is too short to reach the top shelf and can reach only the bottom shelf. Thus, if she comes looking for the sweets, either they will remain on the top shelf or she will take them from the bottom shelf into her room. Figure 1 shows the particular structure of these story worlds.

We can see that in the case of the girl looking for the sweets, the subjunctive past question about the boy, "If not the little girl but the tall boy had come along looking for sweets, where would the sweets be?" gives the correct answer "in the boy's room." This answer would also result from basic condi-

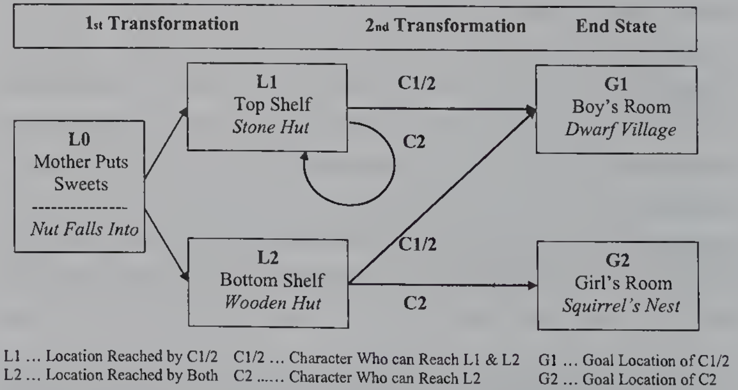


Figure 1. Structure of story worlds in Experiment 1.

tional reasoning: "If 'whenever' the boy comes looking for the sweets, they end up in his room." However, matters are different in the complementary cases in which we ask a subjunctive past question about the girl. Children who are aware of the counterfactual nature of the question—those who know to be selective of the different possible worlds—would answer with "in her room" when sweets had been on the bottom shelf but with "on the top shelf" when sweets had been there. In contrast, children using basic conditional reasoning, despite the subjunctive mood of the question, will give the same answer in both cases, whether the sweets were on the top or the bottom shelf. What their answer will actually be depends on which background assumptions they are more likely to make. If they assume that the sweets are where the girl can reach them, the completion of "If 'whenever' the girl comes looking for sweets, . . ." is ". . . they end up in the girl's room" regardless of where they had been at the beginning. However, when the background assumption is that the sweets are where the girl cannot reach them, then the consequent is that the sweets will be on the top shelf. We do not yet know what background assumptions children are likely to make, as this remains an empirical issue. However, whatever they prefer, as a result of the fixed background assumptions, they should give the same answer in both cases, regardless of where mother puts the sweets. Importantly, this means that we can distinguish between children who employ counterfactual reasoning and those who apply basic conditional reasoning to answer our subjunctive past questions.

Method

Participants. The sample consisted of 33 children (15 boys, 18 girls) from three different nursery schools in the area of Salzburg and Upper Austria, between the age of 2;11 (years; months) and 5;9. The children had a mean age of 4;4 with a standard deviation of 9 months. For the later analysis, the overall sample was split into two age groups. The age of the 16 children in the younger group ranged from 2;11 to 4;4 (M = 3;8, SD = 0;4), and the 17 children in the older group were between 4;5 and 5;9 (M = 4;11, SD = 0;6).

Materials. We used two different wooden models built on 42 × 30 cm wooden platforms. The model for the sweets story consisted of a cupboard with a shelf centrally placed and two boxes, a brown one standing on the shelf and an orange one standing beneath the shelf. For acting out the story, we used sweets and two dolls, a female one and a male one twice as tall as the female one. Each doll had its own room including a table and a photo of itself.

The model for the dwarf story consisted of two huts, one made of wood and one made of stone, and a large walnut tree standing behind the huts. For acting out the story, we used walnuts, a dwarf, and a squirrel. The dwarf lived in the dwarf village (three houses were drawn on a piece of cardboard) and the squirrel lived in a nest in another tree.

Design. We tested all children in two sessions approximately 2 days apart; sessions lasted about 15 min. We administered the two story worlds and the four different event sequences in each story world in fully counterbalanced order, and both story worlds were administered in each session. We tested children with a total of eight indicative future questions and eight subjunctive past questions. In fact, all eight subjunctive past questions required counterfactual reasoning, but children could get away with basic conditional reasoning for six of the subjunctive past questions: In Conditions 2 and 4 (see Table 1) and in either Condition 1 or 3, it is not yet clear which one of the two conditions the children will favor, as this depends on the background assumptions children make.

Table 1

Expected Answers for Each Reasoning Strategy in the Four Conditions of Experiment 1

	Actual events		CF change		Reasoning strategy	
Condition	First transf.	Second transf.	End state	Second transf.	CF	BC
1	L1	C1/2	G1	C1/2 → C2	L1	? (L1, G2)[a]
2	L1	C2	L1	C2 → C1/2	G1	G1
3	L2	C1/2	G1	C1/2 → C2	G2	? (L1, G2)[a]
4	L2	C2	G2	C2 → C1/2	G1	G1

Note. "x → y" indicates that the real event x was replaced by the counterfactually assumed event y, that is, "if not x but y had happened." The original German version for the indicative future question was: "Was passiert mit den Zuckerl wenn jetzt der Simon kommt und nach Zuckerl sucht?" and for the past subjunctive question: "Aber, wenn nicht der Simon sondern die kleine Julia nach Zuckerl gesucht hätte, wo wären denn dann die Zuckerl?" Transf. = transformation; CF = counterfactual; BC = basic conditional; L1 = sweet is on the top shelf/nut falls into the stone hut; L2 = sweet is on the bottom shelf/nut falls into the wooden hut; C1/2 = boy/dwarf comes along; C2 = girl/squirrel comes along; G1 = boy's room/dwarf village; G2 = girl's room/squirrel's nest.

[a] No prediction about which of the two answers children will give, but they should give the same in both conditions.

Procedure. We tested each child in a quiet area away from the other children. We explained the sweets story as follows: These are Simon and his little sister Julia. Both like sweets very much. When their mother buys sweets, she puts them *either* in a box on the top shelf *or* in a box on the bottom shelf. The boy can reach both shelves. Look, when he finds some sweets he brings them into his room. His little sister is not tall enough to reach the top shelf. She can only reach the bottom shelf. If she finds some sweets, she brings them into her room.

The dwarf story is about a dwarf who lives in a dwarf village and a squirrel that lives in a nest in a tree. Both like nuts very much. In search of nuts, they come to a nut tree under which two huts are built. In the roof of each hut there is a tiny hole through which nuts fall inside. The dwarf can open the door of each hut to get in, collect the nuts, and bring them to his village. The squirrel, unfortunately, cannot open the doors of the huts. However, through a hole in the wooden hut, it can squeeze inside to collect the nuts and bring them to its nest.

Children had to answer a total of seven control questions per story world to ensure that they had remembered all details, such as (1) Which one is the boy's room? (2) Which one is the girl's room? (3) From which shelf can the boy take sweets? (4) Where does he bring the sweets? (5) From which shelf can the girl take sweets? (6) Where does she bring the sweets? (7) Why is the girl unable to take the sweets from the top shelf? Examiners corrected wrong answers and familiarized the children by repeating the story (two times at maximum) until they answered all control questions correctly.

After the children became familiar with the structure of a story world, testing started with different events, resulting from the combination of where the object of desire was located (L1, L2) and which character came to collect it (C1/2, the character that can reach each location, and C2, who can reach only Location L2). For each event, we asked indicative future and subjunctive past questions. For instance (Condition 1): "Today the box on the bottom shelf is empty. There are only sweets in the box on the top shelf." At this point, we asked the children the following questions:

> MEMORY 1: "Where are the sweets now?"
>
> INDICATIVE FUTURE: "What will happen to the sweets when the boy comes looking for sweets?"

After the child has given an answer, this event is played out: "Look! This time the boy comes along, looking for sweets. He finds them in the box on the top shelf and takes them to his room!"

> MEMORY 2: "Where are the sweets now?"
>
> SUBJUNCTIVE PAST: "But what if not the tall boy but the little girl had come along looking for sweets, where would the sweets be?"

Table 1 shows the expected answers for each condition depending on whether children use counterfactual or basic conditional reasoning to answer the subjunctive past question.

Results

Control and memory questions. The children answered the different control and memory questions almost perfectly (above 95%) on the first trial. They appeared to have a good understanding of how the story structure worked. We checked that the exclusion from our analysis of children who answered incorrectly had no influence on our results.

Indicative future questions. Performance on the indicative future questions of the sweets story (94% correct) and dwarf story (96% correct) was not significantly different, Wilcoxon signed ranks test: $Z(N = 33) = -1.0$, $p = .32$, Cohen's $d = 0.17$. Eleven children made one mistake (out of eight questions); thus, they answered 253 of 264 questions (almost 96%) correctly. However, there was a significant difference between the four conditions (see Table 1) that asked indicative future questions, Friedman test: $\chi^2(3,\ N = 33) = 27.4$, $p < .01$, $\eta_p^2 = .28$. The children made 10 mistakes in Condition 2, where the character C2 could not reach the item in Location L1, such as when mother had put the sweets on the top shelf: "What will happen with the sweets now, when the little girl comes looking for sweets?" The younger children in particular had some difficulty in understanding that the sweets would stay on the top shelf. Nevertheless, they answered 75% of the indicative future questions in Condition 2 correctly—every child answered at least one of the two questions correctly—compared to the older group, who answered 87.5% of the questions correctly. Clearly, most of the younger and practically all of the older children understood that the item stays in L1 when the C2 character comes. Children who answered only one of these indicative future questions correctly did not perform significantly differently on the corresponding subjunctive past question (the tall boy had come to take the sweets from the top shelf L1 to his room: "What if not the tall boy but the little girl had come along, where would the sweets be?"—in L1) from those children who answered both of the indicative future questions correctly, Mann–Whitney test: $U(n_1 = 10, n_2 = 23) = 100$, $p = .24$, Cohen's $d = 0.43$. We checked that exclusion of the 10 children who answered the indicative future question of Condition 2 incorrectly had no influence on the interpretation of results.

Subjunctive past questions. Children had noticeably more problems with the subjunctive past questions, where they gave correct answers to 55% of the questions in the sweets story and 46% in the dwarf story, Wilcoxon signed ranks test: $Z(N = 33) = -1.5$, $p = .13$, Cohen's $d = 0.28$. Thus, they answered only 50% correctly compared to 96% of the indicative future questions, Wilcoxon signed ranks test: $Z(N = 33) = -5.0$, $p < .01$, Cohen's $d = 2.34$. However, performance is above chance if we consider the chance level to be at 25% (four possible answers: L1, L2, G1, and G2). The older children gave more (58.8%) correct answers than the younger children (42.1%), Mann–Whitney test: $U(n_1 = 16, n_2 = 17) = 63$, $p < .01$, Cohen's $d = 0.89$. Figure 2 shows children's performance on the subjunctive past questions, split up by the four different conditions (averaged over the two story worlds).

We found that particular conditions had significant effects in the subjunctive past questions, Friedman test: $\chi^2(3, N = 33) = 53.9$, $p < .01$, $\eta_p^2 = .55$. Condition 1 differed significantly from all the other conditions, Wilcoxon signed ranks test for Condition 2: $Z(N = 33) = -4.8$, $p < .01$, Cohen's $d = 1.81$; for Condition 3: $Z(N = 33) = -4.4$, $p < .01$, Cohen's $d = 1.28$; for Condition 4: $Z(N = 33) = -4.6$, $p < .01$, Cohen's $d = 1.54$; but Conditions 2, 3, and 4 did not differ significantly from each other, Wilcoxon signed ranks test for Conditions 2 and 3: $Z(N = 33) = -1.8$, $p = .07$, Cohen's $d = 0.31$; for Conditions 2 and 4: $Z(N = 33) = -1.0$, $p = .3$, Cohen's $d = 0.17$; for Condition 3 and 4: $Z(N = 33) = -1.2$, $p = .3$, Cohen's $d = 0.20$.

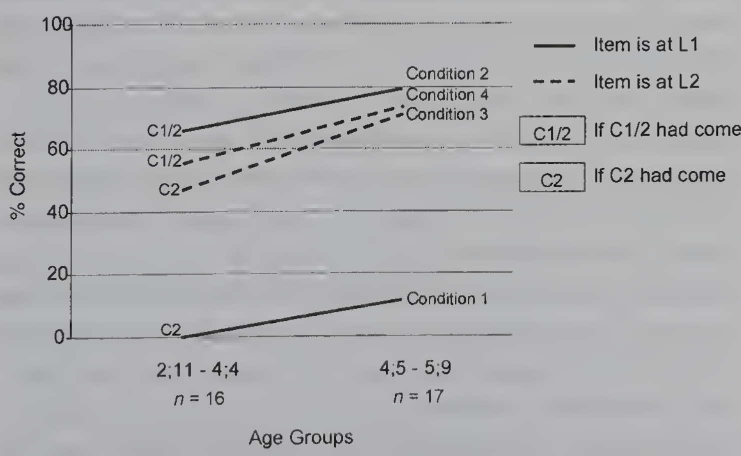


Figure 2. Counterfactual reasoning performance on the four different conditions in Experiment 1.

Our results show that children had severe difficulties, especially with subjunctive past questions where the C2 character could not reach the sweets at L1 (Condition 1): For example, mother has put the sweets in the top shelf (L1), and the boy came and took them to his room (G1): "If not the boy but the little girl had come along looking for sweets, where would the sweets be?"—in L1. The children answered only 4 of 66 questions (6%) correctly, which is clearly below chance. The common error—as predicted if children engage in basic conditional reasoning and use the default assumption that each character takes sweets to his or her room—was to say that sweets would have ended up in the C2 character's room, Location G2 (39 of 66 questions = 59%; 64% of these children had answered with G2 in Condition 3 too). The children answered 10 of 66 questions (15%) with the actual location (G1: realist error) and 13 questions (20%) with the bottom shelf (L2).

In comparison, children gave 40 correct answers to 66 subjunctive past questions (61%) when the C2 character could have reached the sweets (Condition 3). This is not surprising if we assume that they applied basic conditional reasoning with a default assumption that sweets tend to end up in the respective character's room. When we asked, "But what if not the little girl but the tall boy (character C1/2) had come along looking for sweets . . .?" we expected most children to answer that the sweets would then be in the boy's room (Location G1) regardless of where mother had put them, and 73% in Condition 2 and 65% in Condition 4 gave the expected G1 answer.

To summarize, children responded correctly above chance levels in Conditions 2, 3, and 4 but below chance in Condition 1. This pattern of results conforms strongly to what we expected under the assumption that children solve problems by basic conditional reasoning (Table 1). To provide evidence of the participants' limited counterfactual reasoning skills, despite their knowledge of how the relevant transformations work, we compared the answers to subjunctive past questions, for example in Condition 1 with the answers to the corresponding indicative future questions, in this case in Condition 2. All four comparisons were highly significantly different (Wilcoxon signed ranks test: Z scores = −3.7 to −5.1; Cohen's ds = 0.83 to 2.81). It seems that the children generally understood the constraint on the girl (which we tested by indicative future, control, and memory questions), but were unable to create the correct counterfactual assumptions to infer that if C2 had come she would not be able to retrieve the sweets from L1.

Discussion

Experiment 1 demonstrates two clear effects. One of these is the replication of the difference between the number of correct answers on indicative future questions and on subjunctive past questions also found in studies by Perner et al. (2004), Riggs and Peterson (2000), Riggs et al. (1998), and Robinson and Beck (2000, Study 1). The other clear effect is that performance on the subjunctive past questions can be very different within the same story world. Children give mostly correct answers if they can derive them by basic conditional reasoning from preferred default assumptions.

In Condition 1 (top shelf, boy comes, takes sweets: If the little girl had come . . .?), the preferred default assumption leads to a wrong answer (girl's room). To get it right, children have to consider where mother had actually put the sweets. In other words, they have to construct a possible world that is maximally similar to the actual world, one that takes account of where the sweets had actually been put. Most children in Experiment 1 failed to do this, in apparent violation of Lewis's (1973) "nearest possible world" criterion.

Experiment 2

The results of Experiment 1 showed that most children younger than 6 did not use counterfactual reasoning but used basic conditional reasoning aided by background assumptions suitable for three but not all four conditions. One worry, however, is that children capable of counterfactual reasoning might have been trapped into unthinkingly using the simplified rule that the sweets end up in the searching character's room because this is the case in three of four event sequences. This is due to an asymmetry in the structure of the story worlds of Experiment 1 (see Figure 1). There are three event sequences in which the item is transferred to the protagonist's room (Locations G1 and G2) but just one event sequence in which the item stays in its initial location (Location L1), and this condition caused children-specific difficulties when we asked them counterfactually in Experiment 1. To eliminate any possibility that children might perform badly because the one condition was an exception, we modified the structure of the story worlds for Experiment 2 so that it was symmetrical. Thus, the structure of Figure 1 changed so that the line labeled C1/2 connecting L2 with G1 now also becomes self-reflective on L2; for example, the boy is no longer able to reach the sweets when they are at the bottom shelf, and therefore the sweets stay at L2. Because he can reach only L1, the C1/2 character becomes a C1 character.

If children's basic conditional reasoning and their preferred background assumptions led to the many errors in Condition 1 of Experiment 1 (where the counterfactual character could not reach the sweets), then in this experiment, children should commit most of their errors in Conditions 1 and 4. However, if children reasoned counterfactually in Experiment 1 but were misled by the fact that the sweets ended up in the character's room in three of four event sequences, then their errors in Experiment 2 should be equally distributed over the four conditions because now the sweets remain on the shelves and are brought to the characters' rooms an equal number of times.

There is a further reason for running this experiment. Children's problems with Condition 1 in Experiment 1 indicated that they apparently disregarded the first transformation (where mother put the sweets) when asked a subjunctive past question about the second transformation (if the little girl had come). This might have happened for different reasons. Children may simply have had difficulty in remembering where mother had put the sweets. Although we could show in Experiment 1 that children had few problems answering indicative future, memory, and control questions—implying that they knew the rules of the story world—it might be that they were unable to retrieve this information when we asked the subjunctive past question. Conversely, children may simply have failed to look further into the past than the counterfactual assumption led them to do. For instance, one can imagine that children, when asked what would have happened if the girl instead of the boy had come, would focus on this point in the story and the ensuing consequences of this counterfactual assumption, and ignore the facts earlier in the story, that is, whether the sweets had been on the top or the bottom shelf. A third possibility is that children did not understand that a subjunctive past question requires them to think of the actual events in the story world to arrive at a correct answer. If the first or second factor was the reason that prevented children from reasoning counterfactually, then making the earlier transformation (mother puts sweets into L1 or L2) counterfactual should solve their problems. So, for example, when we ask "But what if sweets had been on the bottom instead of the top shelf?" children need not look further into the past; moreover, the current location of the sweets reminds them of who had come to collect the sweets. However, if the third factor was the reason for giving the wrong answer, then the children should now have problems with Conditions 1 and 4 even when the first transformation is taken counterfactually. To test this, we introduced subjunctive past questions about the first transformation, where mother had put the sweets.

Method

Participants. Thirty-two children (19 boys, 13 girls) from two different preschools and nursery schools in Upper Austria, between the age of 5;0 and 6;5, participated in this study. They had a mean age of 5;10 with a standard deviation of 4 months. All participants were monolingual German speakers and came from working- and middle-class backgrounds. For the later analysis, the overall sample was split into two age groups. The 15 younger children were between the ages of 5;0 and 5;9 ($M = 5;6$, $SD = 0;2$), and the 17 children in the older age group were between 5;10 and 6;5 ($M = 6;1$, $SD = 0;2$).

Materials. We used the same materials as in Experiment 1, except that the dwarf story contained only the stone hut, not the wooden hut.

Design. We tested all children in one 20-min session during which we presented both story worlds (sweets story and dwarf story) and asked children eight control questions per story world. We used the same seven as in Experiment 1, with an additional question: "Why is the boy unable to take the sweets from the bottom shelf?" We gave each child four event sequences, two per story world, and we asked two test questions per sequence: an indicative future question (the same as in Experiment 1) and a subjunctive past question about either the first ("What if sweets had not been at L1 but at L2?") or the second transformation ("What if not C1 but C2 had come?"). Thus, each child answered four subjunctive past questions, two in which the location was the counterfactual element (for one story world the character could have reached the item, and for the other story world the character could not), and two in which the character was the counterfactual element (for one story world the character could have reached the item, and for the other story world the character could not). We used a Latin square design to control for variation.

Procedure. The procedure was basically the same as in Experiment 1, save for two modifications. In the sweets story, the tall boy had broken his leg and could not kneel down to reach the sweets on the bottom shelf. In the dwarf story, there was only one hut, the stone one whose door the dwarf can open to collect nuts. The squirrel cannot open the door of the stone hut but can climb up the tree to get to the nuts, whereas the dwarf is overweight and is unable to climb the tree. All nuts are *either* in the stone hut *or* on the tree.

Results

Control and memory questions. Children's performance in the control and memory questions of the sweets story and the dwarf story was excellent and practically identical for both story worlds, 98.4% versus 99.2%.

Indicative future questions. Children's performance on the indicative future questions was very good and did not differ significantly between the two story worlds (sweets story: 100% correct; dwarf story: 91% correct), Wilcoxon signed ranks test: $Z(N = 32) = -1.9$, $p = .06$, Cohen's $d = 0.35$. Two children made one mistake out of four questions, whereas 2 children made two mistakes; thus, overall, children answered almost 96% of the questions (122 of 128) correctly. Children did not perform significantly differently on indicative future questions, where the character is or is not able to reach the item, Wilcoxon signed ranks test: $Z(N = 32) = 0.0$, $p = 1.0$. Children seemed to understand that in some scenarios, the item stays where it is; only the youngest age group in Experiment 1 had problems with the indicative future questions where the item stayed in its location.

Subjunctive past questions. Children gave 59% correct answers to the subjunctive past questions in the sweets story and 75% in the dwarf story, Wilcoxon signed ranks test: $Z(N = 32) = -1.8$, $p = .08$, Cohen's $d = 0.31$. The Wilcoxon signed rank test showed a highly significant difference between answers to indicative future (96% correct) and to subjunctive past questions (63%), $Z(N = 32) = -4.5$, $p < .01$, Cohen's $d = 1.41$. Improvement on subjunctive past questions with age from 63% to 71%

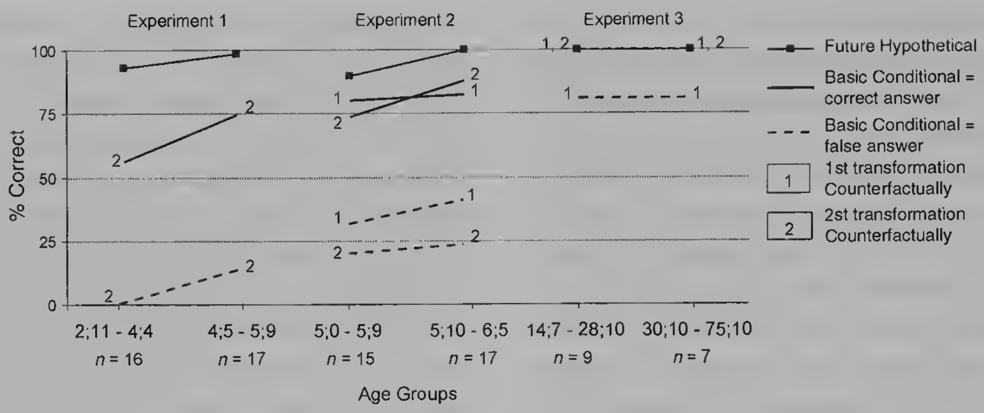


Figure 3. Percentage of correct answers to counterfactual questions depending on age.

correct was minimal and not significant, Mann–Whitney test: $U(n_1 = 15, n_2 = 17) = 103$, $p = .32$, Cohen's $d = 0.32$.

Figure 3 shows that Experiment 2 replicates the results of Experiment 1 rather closely: When the second transformation (searching character) is taken counterfactually, children do significantly worse when the character would not have reached the item (21.9% correct; dashed line, Conditions 1 and 4) than when he or she would have reached it (81.3% correct; solid line, Conditions 2 and 3), Wilcoxon signed ranks test: $Z(N = 32) = -4.2$, $p < .01$, Cohen's $d = 1.06$. There were no significant differences between Conditions 1 and 4, Mann–Whitney test: $U(n_1 = 16, n_2 = 16) = 104$, $p = .21$, Cohen's $d = 0.47$; or Conditions 2 and 3, Mann–Whitney test: $U(n_1 = 16, n_2 = 16) = 128$, $p = 1.0$. Thus, children's errors in Condition 1 of Experiment 1 did not result simply from the fact that it was the only condition in which the protagonist cannot reach the item.

The novel part of Experiment 2 was that we also asked children subjunctive past questions that altered the first transformation (where mother put sweets). Response patterns were similar (see Figure 3) and not significantly different from performance on the second transformation, Wilcoxon signed ranks test for first and second transformations: Conditions 1 and 4: $Z(N = 32) = -1.3$, $p = .20$, Cohen's $d = 0.35$; Conditions 2 and 3: $Z(N = 32) = 0.0$, $p = 1.0$. Children did significantly worse when the item would have stayed at its initial location (37.5% correct, dashed line, Conditions 1 and 4) than when it would have ended up in the character's goal location (81.3% correct, solid line, Conditions 2 and 3), Wilcoxon signed ranks test: $Z(N = 32) = -3.3$, $p < .01$, Cohen's $d = 0.71$. Neither Conditions 1 and 4, Mann–Whitney test: $U(n_1 = 16, n_2 = 16) = 96$, $p = .08$, Cohen's $d = 0.68$; nor Conditions 2 and 3, Mann–Whitney test: $U(n_1 = 16, n_2 = 16) = 120$, $p = .63$; Cohen's $d = 0.17$, showed a significant difference between the number of correct answers for the first or second transformation.

The pattern of answers on our subjunctive past questions allowed us to classify children according to the reasoning strategy that would result in the observed pattern. Table 2 shows the response patterns expected for counterfactual and basic conditional reasoning strategies. As children did not receive all eight conditions but only two for each transformation, the table shows which combination of conditions we gave to one group (a) and to the other group (b).

We classified children as *counterfactual reasoners* (CF) if they showed the counterfactual reasoning pattern on all four conditions (2 children) or on at least three conditions (4 children, allowing one lapse of attention). Four older (23%) and 2 younger (13%) children showed a counterfactual reasoning strategy. *Basic conditional reasoners* (BC) were those who showed the basic conditional reasoning strategy on all four (11 children) or at least three conditions (1 child), which applied to 6 older (35%) and 6 younger (40%) children. The binomial probability for inclusion in the CF group is as high as for the BC group: .02. An additional 4 children, 2 older (12%) and 2 younger (13%), showed a particular kind of incomplete basic conditional reasoning pattern, which resulted when they switched the person who came to search for the sweets (*searcher switchers* [SS]) when the location was counterfactually altered. For example, if the sweets are at L2, the C2 character comes searching for sweets, they end up in C2's room, and we ask children, "Where would the sweets be, if they had not been at L2 but at L1?" children pointed to C1's room. Although this pattern is technically BC (with one very specific lapse, which has a probability of .01), we kept it separate because, as we will see, we found the same pattern among some adults who otherwise

Table 2

Expected Answers for Each Reasoning Strategy in the Eight Conditions of Experiment 2

Condition	Actual events			CF change First transf.	Reasoning strategy		CF change Second transf.	Reasoning strategy	
	First transf.	Second transf.	End state		CF	BC		CF	BC
1	L1	C1	G1 a	L1 → L2	L2	G1 b	C1 → C2	L1	G2
2	L1	C2	L1 a	L1 → L2	G2	G2 b	C2 → C1	G1	G1
3	L2	C1	L2 b	L2 → L1	G1	G1 a	C1 → C2	G2	G2
4	L2	C2	G2 b	L2 → L1	L1	G2 a	C2 → C1	L2	G1

Note. "x → y" indicates that the real event x was replaced by the counterfactually assumed event y, that is, "if not x but y had happened." Transf. = transformation; CF = counterfactual; BC = basic conditional; L1 = sweet is on the top shelf/nut falls into the stone hut; L2 = sweet is on the bottom shelf/nut is on the nut-tree; C1 = boy/dwarf comes along; C2 = girl/squirrel comes along; G1 = boy's room/dwarf village; G2 = girl's room/squirrel's nest.

showed only counterfactual patterns. There were 3 children, 2 older (12%) and 1 younger (7%), whose pattern could be equally correctly classified as counterfactual or basic conditional reasoning (with one lapse of attention, CF/BC, with a probability of .008). Finally, the responses of 7 children, 3 older (18%) and 4 younger (27%), did not fit any of these response patterns, and we classified them as *nonsystematic.*

Discussion

We replicated the findings that children were much better at answering indicative future questions than subjunctive past questions in Experiment 2. There is clear evidence that children find future hypothetical reasoning much easier than counterfactual reasoning. Another aim of this experiment was to exclude the possibility that children's errors in Condition 1 of Experiment 1 occurred because it was the only condition in which the little girl could not have retrieved the sweets from the top shelf. The second experiment reveals that this was not the case. Despite balanced numbers, children still had problems in the conditions (1 and 4) in which basic conditional reasoning leads to a different answer than counterfactual reasoning does.

We also wanted to test alternative reasons for why children had problems in Conditions 1 and 4, specifically, difficulty in remembering the original location of the item and not looking farther into the past than the counterfactual assumption takes them. If these alternative accounts applied, then children should have found it easier to draw correct conclusions when we made the first transformation counterfactual than when the second was counterfactual. However, this was not the case. Children most likely used basic conditional reasoning rather than counterfactual reasoning.

This conclusion is underlined by the fact that, similar to the results of Experiment 1, at most 3 of the younger and 6 of the older children followed a consistent counterfactual reasoning strategy (also counting the CF/BC patterns as counterfactual) when answering subjunctive past questions, whereas 8 of the younger and 8 of the older children tended to display basic conditional reasoning (also counting the SS group, which showed a one-lapse basic conditional reasoning pattern).

Experiment 3

In particular, the results of the last experiment indicated that even at 6 years, few children, at best 35% (if the ambiguous CF/BC pattern is counted as CF), showed evidence of counterfactual reasoning. We wondered whether adults might be less than perfect on our problems, so in this experiment, we told the stories of Experiment 2 to a group of adults.

Method

Participants. Sixteen adults (7 men, 9 women) from ages 14;7 to 75;10 participated in this study. The youngest attended secondary school, 9 were university students enrolled in different fields of study, and the remaining 6 were middle-class employees. They had a mean age of 34;6 with a standard deviation of 16;3.

Procedure. The procedure was exactly the same as in Experiment 2.

Results and Discussion

The adults answered all control, memory, and indicative future questions correctly. Moreover, as Figure 3 shows, they also answered all subjunctive past questions correctly with one exception. When the first transformation (where mother put the sweets) was taken counterfactually, they gave only 81% correct answers, because 6 people showed the SS pattern we described for the small group of children in Experiment 2. When the object was counterfactually assumed to have been in a location where the actual character could not reach it, the participants assumed the character who could reach there would eventually come and fetch it.

At first blush, this appears to be a default assumption that 4 children also tended to make. However, unlike the 4 children, adults explained that they did not only reason about the fact that, for example, the girl cannot reach the sweets on the top shelf and thus sweets would have stayed there, but they continued to reason that at some later time the boy would come and take the sweets into his room. From these explicit explanations, it seems clear that adults did not engage in basic conditional reasoning that ignored the actual events. Another fact in favor of adults' counterfactual reasoning is that none of them showed the 4 children's tendency to use basic conditional reasoning when the second transfer was taken counterfactually.

As we did not elicit explanations from children about their answers, we cannot tell whether they did or did not follow the same strategy as adults. However, the 4 children also answered with the character's goal locations when the second transformation was counterfactually changed, which adults did not do, so the more parsimonious explanation is that the children showing the SS pattern took a basic conditional approach with the default assumption that the characters tend to bring the goods to their room. This can account for their answer patterns in both cases, when the first and the second transformations were taken counterfactually.

General Discussion

The principal aim of these studies was to get a clearer picture of children's ability to reason counterfactually by controlling for the distinction between counterfactual and other conditional reasoning strategies. With this distinction, we also hoped to shed light on the discrepancy between previous studies' claims about the onset of counterfactual reasoning.

In short, our explanation for this discrepancy is that previous research did not control for children's use of basic conditional reasoning strategies to answer subjunctive questions about the past. Counterfactual problems seem easy if one can correctly answer them with purely conditional reasoning (e.g., at 3 years, Harris et al., 1996; or 4 years, Riggs et al., 1998). They become very difficult when basic conditional reasoning with plausible default assumptions produces different answers than proper counterfactual reasoning would produce, as this research shows. Indications of that difficulty are also visible in children's understanding of counterfactual emotions like regret or relief (Amsel & Smalley, 2000; Amsel et al., 2003). To assess whether children reason counterfactually by following the ''nearest possible world'' constraint of Lewis (1973), we used counterfactual problems for which basic conditional reasoning would not produce correct answers. These problems required children and adults to think of what has actually happened and constrain the necessary assumptions for conditional reasoning to the actual events as far as logically possible (known as the nearest possible world). Results show that children, but not adults, tend to follow default assumptions unconstrained by actual events and governed by cognitive availability.

There are, however, competing explanations for results showing discrepancies in the age of onset of counterfactual reasoning. One claim is that children are more likely to engage in counterfactual thinking when the outcome is unwanted or negative (German, 1999; Guttentag & Ferrell, 2004; Roese, 1997). This factor could have contributed to our findings, as children had the hardest time with conditions in which a character reached the desired object, whereas they had an easier time with conditions in which the character failed to reach the desired object (negative outcomes supposedly stimulate counterfactual reasoning). However, as we discussed in the Introduction, many earlier studies found that negative outcomes had only minor effects on counterfactual reasoning (Amsel & Smalley, 2000; Amsel et al., 2003) or no advantage at all (Harris, 2000; Kavanaugh & Harris, 2000), so it is unlikely that the massive difference we found in the present study could be due to this possible but minor effect.

Another possible explanation for children's problems with subjunctive past questions is that they have difficulty in inhibiting their strong initial responses based on the current reality (Riggs et al., 1998). Beck et al. (2009) found that inhibitory control predicted children's performance on the counterfactual reasoning tasks; in contrast, Robinson and Beck (2000) questioned whether inhibition is a satisfactory explanation, as in many counterfactual reasoning tasks the only error children can plausibly make is to choose the current state of affairs. So, it is unclear whether children do in fact fail to inhibit reality or whether they fail to construct a counterfactual alternative. Robinson and Beck reasoned further that if children do tend to get trapped in current reality, then it should be easier for them to think of a counterfactual alternative to a *past* reality than to a *current* reality. Children in their study did equally well in both conditions, suggesting that they do not get trapped in current reality. In a similar vein, Riggs et al. (1998) pointed out that inhibitory control cannot account for the difference in performance between counterfactual and future hypothetical reasoning because otherwise children should have problems inhibiting reality in most future hypothetical reasoning tasks as well, as the hypothetically assumed antecedent often differs from known reality, but in fact they do not experience difficulty with future hypotheticals. In any case, in our stories children hardly suffered from realist errors. Their typical errors consisted of answering with another counterfactual location, not with the object's current location. Hence, our results are clearly not the result of weak executive control in inhibiting realist answers.

Working memory limitations, in contrast, might provide a more convincing account for children's failure to engage in counterfactual reasoning. One could argue that children might understand that they should construct a nearest possible world by taking actual events into account but fail to do so because they lack the memory capacity to remember the actual event, construct the counterfactual scenario, and integrate that counterfactual with the actual event. Beck et al. (2009) tested whether working memory capacity can predict 4-year-old children's success on counterfactual reasoning tasks. However, the relation between working memory tasks (Counting and Labeling; Noisy Book) and counterfactual reasoning tasks (causal chains by German & Nichols, 2003; unexpected transfer by Riggs et al., 1998, and false syllogisms) was at most .28. Thus, an overly taxed working memory cannot exclusively—if at all—explain children's difficulties with counterfactual reasoning tasks. Working memory might explain younger children's difficulty with subjunctive past questions when holding two models simultaneously active, but it cannot explain the older children's problems.

Thus, our central finding is that children only gain the ability to reason counterfactually as late as 5 or 6 years. This late onset might also explain why Amsel et al. (2003; Amsel & Smalley, 2000) as well as Guttentag and Ferrell (2004) report such late understanding of counterfactual emotions. We designed our tasks so that in the critical conditions children could only give correct answers if they naturally adhered to the "nearest possible world" constraint, that is, if they related their conditional reasoning to actual events. This makes it similar to showing counterfactual emotions, except that the direction of the relation is reversed. For instance, regret can be felt only if one compares what one actually received with what one could have received if one had chosen or acted differently. If children up to 6 years or older do not spontaneously compare assumed conditions with actual conditions, they would perform badly on our tasks and on the counterfactual emotions tasks.

On some other counterfactual tasks (defined by the use of past subjunctive questions), however, much younger children before the age of 5 years perform practically at ceiling (Harris et al., 1996; Riggs et al., 1998). We can explain this by assuming that in these tasks, basic conditional reasoning leads to the same "correct" answers as counterfactual reasoning.

Our suggestion that children give correct answers based on basic conditional reasoning opens up a new venue of explanation. It is related to the idea that with each additional step, it becomes less likely that all of the required default assumptions match the "nearest possible world" criterion. For instance, in the scenarios by Harris et al. (1996), the stipulated antecedent "If Carol had taken her shoes off, would the floor be dirty?" triggers fairly directly the relevant knowledge that without shoes, the floors stay clean. In the scenario Riggs et al. (1998, Experiment 1) used, the reasoning is not so straightforward: "If there had been no fire, where would Peter be?" triggers knowledge that without a fire, firemen could be anywhere. However, the story provides children with two possible locations, namely, the ones mentioned in conjunction with Peter: home or post office. With the reality bias diminishing, children will increasingly bring more sophisticated assumptions to bear, for example, that people are more likely to be at home than at the burning post office. Such knowledge would increase "home" responses even though children are still applying basic conditional reasoning.

The findings by Amsel, Trionfi, and Campbell (2005) seem to support this assumption. They compared 6-year-olds, 10-year-olds, and college students' performances on two conditions: (a) *pretend* that in a make-believe world dogs meow and (b) *imagine* what the real world would be like if dogs meowed. Both conditions presented a minor premise, "Rover is a dog," followed by a question, "Does Rover meow?" The main finding was that 6- and 10-year-old children performed correctly (saying *yes*) more frequently in the pretend condition than in the imagine condition, whereas this difference vanished among college students. The crucial difference is that in the pretend condition, one does not assume that dogs meow in the real world, whereas in the imagine condition one must assume this, which obviously makes the imagine condition a counterfactual, as dogs do not meow in reality. It is well known that children perform better on counterfactual syllogism tasks when the counterfactual premise is framed as pretend (Dias & Harris, 1988; Richards & Sanderson, 1999). This is especially interesting when we consider that the children in our studies found it easier to make background assumptions that were not constrained by reality than to reason with assumptions that were constrained by reality.

The assumption that children use basic conditional reasoning in counterfactual tasks is compatible with a study by Robinson, Rowley, Beck, Carroll, and Apperly (2006), who used a setup in which a mouse went down a slide that split halfway down into a spotted or a striped slide. In one condition, researchers asked the questions when it was still unknown which way the mouse would go (physical possibilities), whereas in the other condition the mouse had already gone into one branch but the child had no knowledge which branch that was (epistemic possibilities). To do the task right, children had to put out two mats to make sure that the mouse that came down a slide did not get hurt. A total of 87% of 6-year-olds put both mats out in the physical-possibilities case, but only 43% did so in the epistemic-possibilities case. The epistemic case shares with counterfactual reasoning the need to construct alternative models for an existing state of the world, an ability that seems to emerge around 6 years of age.

Conclusion

We found a large discrepancy in correct answers to past subjunctive questions under two different conditions. When basic conditional reasoning yielded the correct answer, children did much better than when they had to observe the "nearest possible world" constraint—that is, when they had to demonstrate proper counterfactual reasoning. With this discovery and known observations from the literature, we can describe children's progress on counterfactual reasoning problems that use past subjunctive questions as follows. At 3 years, children still have a noticeable reality bias, but this bias practically disappears by 4.5 years. Children at this age are, however, still not perfect on counterfactual problems because they use basic conditional reasoning instead of counterfactual reasoning. The likelihood of a correct answer depends on what general knowledge they can bring to bear on the conditional antecedent presented in the question and how directly it indicates the correct answer, that is, whether it leads specifically to the correct answer or only contains that answer in a set of options. By 5 or 6 years, some children start to reason counterfactually. They do not leave their reasoning exposed to the vacillations of cognitive accessibility; instead, they understand that it must follow the givens of the actual events as closely as possible (the nearest possible world). Adults' mastery is close to perfection, at least on our kinds of task. However, one could argue that the few older children who give "counterfactual" answers, instead of using counterfactual reasoning, might have used basic conditional reasoning together with expedient background assumptions—in our study, that the sweets were stored in the top or bottom shelves as a result of where the sweets had actually been stored. This remains a possibility, which future research should clarify. We also leave it to future research to plot the developmental improvements from 6 years to adulthood.

References

Amsel, E., Robbins, M., Tumarkin, T., Janit, A., Foulkes, S., & Smalley, J. D. (2003). *The card not chosen: The development of judgments of regret in self and others*. Unpublished manuscript.

Amsel, E., & Smalley, J. D. (2000). Beyond really and truly: Children's counterfactual thinking about pretend and possible worlds. In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning and the mind* (pp. 121–147). Hove, UK: Psychology Press.

Amsel, E., Trionfi, G., & Campbell, R. (2005). Reasoning about make-believe and hypothetical suppositions: Towards a theory of belief-contravening reasoning. *Cognitive Development*, *20*, 545–575.

Beck, S. R., Riggs, K. J., & Gorniak, S. L. (2009). Relating developments in children's counterfactual thinking and executive functions. *Thinking and Reasoning*, *15*, 337–354.

Beck, S. R., Robinson, E. J., Carroll, D. J., & Apperly, I. A. (2006). Children's thinking about counterfactuals and future hypotheticals as possibilities. *Child Development, 77,* 413–426.

Byrne, R. M. J. (2005). *The rational imagination: How people create alternatives to reality*. Cambridge, MA: MIT Press.

Dias, M. G., & Harris, P. L. (1988). The effect of make-believe play on deductive reasoning. *British Journal of Developmental Psychology, 6,* 207–221.

German, T. P. (1999). Children's causal reasoning: Counterfactual thinking occurs for ''negative'' outcomes only. *Developmental Science, 2,* 442–447.

German, T. P., & Nichols, S. (2003). Children's counterfactual inferences about long and short causal chains. *Developmental Science, 6,* 514–523.

Grant, C. M., Riggs, K. J., & Boucher, J. (2004). Counterfactual and mental state reasoning in children with autism. *Journal of Autism and Developmental Disorders, 34,* 177–188.

Guajardo, N. R., & Turley-Ames, K. J. (2004). Prescholers' generation of different types of counterfactual statements and theory of mind understanding. *Cognitive Development, 19,* 53–80.

Guttentag, R., & Ferrell, J. (2004). Reality compared with its alternatives: Age differences in judgments of regret and relief. *Developmental Psychology, 40,* 764–775.

Harris, P. L. (2000). *The work of the imagination: Understanding children's worlds*. Oxford, UK: Blackwell.

Harris, P. L., German, T., & Mills, P. (1996). Children's use of counterfactual thinking in causal reasoning. *Cognition, 61,* 233–259.

Kavanaugh, R. D., & Harris, P. L. (2000). *Contrasting what did and did not happen: Children's sensitivity to casually relevant contrasts*. Unpublished manuscript.

Lewis, D. (1973). *Counterfactuals*. Oxford, UK: Blackwell.

Markovits, H., & Barrouillet, P. (2002). The development of conditional reasoning: A mental model account. *Developmental Review, 22,* 5–36.

Müller, U., Miller, M. R., Michalczyk, K., & Karapinka, A. (2007). False belief understanding: The influence of person, grammatical mood, counterfactual reasoning and working memory. *British Journal of Developmental Psychology, 25,* 615–632.

Perner, J. (2000). About + belief + counterfactual. In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning and the mind* (pp. 367–401). Hove, UK: Psychology Press.

Perner, J., & Rafetseder, E. (in press). Hypothetical reasoning or counterfactual reasoning: Children lost in the nearest possible world. In C. Hoerl, T. McCormack, & S. Beck (Eds.), *Understanding counterfactuals/understanding causation.* Oxford, UK: Oxford University Press.

Perner, J., Sprung, M., & Steinkogler, B. (2004). Counterfactual conditionals and false belief: A developmental dissociation. *Cognitive Development, 19,* 179–201.

Richards, C. A., & Sanderson, J. A. (1999). The role of imagination in facilitating deductive reasoning in 2-, 3- and 4-year olds. *Cognition, 72,* B1–B9.

Riggs, K. J., & Peterson, D. M. (2000). Counterfactual thinking in pre-school children: Mental state and causal inferences. In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning and the mind* (pp. 87–99). Hove, UK: Psychology Press.

Riggs, K. J., Peterson, D. M., Robinson, E. J., & Mitchell, P. (1998). Are errors in false belief tasks symptomatic of a broader difficulty with counterfactuality? *Cognitive Development, 13,* 73–90.

Robinson, E. J., & Beck, S. (2000). What is difficult about counterfactual reasoning? In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning and the mind* (pp. 101–119). Hove, UK: Psychology Press.

Robinson, E. J., Rowley, M. G., Beck, S. R., Carroll, D. J., & Apperly, I. A. (2006). Children's sensitivity to their own relative ignorance: Handling of possibilities under epistemic and physical uncertainty. *Child Development, 77,* 1642–1655.

Roese, N. J. (1997). Counterfactual thinking. *Psychological Bulletin, 121,* 133–148.

Rumain, B., Connell, J., & Brain, M. D. S. (1983). Conversational comprehension processes are responsible for reasoning fallacies in children as well as adults: If is not the biconditional. *Developmental Psychology, 19,* 471–481.

Child Development, January/February 2010, Volume 81, Number 1, Pages 390–409

Sensing the Coherence of Biology in Contrast to Psychology: Young Children's Use of Causal Relations to Distinguish Two Foundational Domains

Jane E. Erickson, Frank C. Keil, and Kristi L. Lockhart
Yale University

To what extent do children understand that biological processes fall into 1 coherent domain unified by distinct causal principles? In Experiments 1 and 2 ($N = 125$) kindergartners are given triads of biological and psychological processes and asked to identify which 2 members of the triad belong together. Results show that 5-year-olds correctly cluster biological processes and separate them from psychological ones. Experiments 3 and 4 ($N = 64$) examine whether or not children make this distinction because they understand that biological and psychological processes operate according to fundamentally different causal mechanisms. The results suggest that 5-year-olds do possess this understanding, and furthermore, they have intuitions about the nature of these different mechanisms.

Young children seem to have a clear sense that there are fundamentally different kinds of things in the world, or ontological categories, and they distinguish them by using different causal-explanatory frameworks (Carey, 1985; Keil, 1989; Murphy & Medin, 1985; Wellman & Gelman, 1992, 1998). Many questions remain, however, about the nature of these causal frameworks and how they guide acquisition and use of knowledge. Related questions ask how these frameworks develop into more sophisticated domain-specific causal ''theories'' possessed by lay adults. Here we explore aspects of the internal structure of such frameworks and consider the implications for models of conceptual development.

Three sorts of causal-explanatory frameworks have dominated much of the discussion on domain-specific causal principles—those fundamental to the domains of physics, psychology, and biology. The discussion revolves around the process through which children become aware of which entities belong in each of these domains and the extent to which the assignment of entities to domains is based on knowledge of deeper causal relations. Even preverbal infants consistently differentiate between animate and inanimate things and display some understanding of the causal principles driving the actions of entities within each of these domains (Baillargeon, 1994; Bertenthal, 1993; Bullock, 1985; Gelman, Durgin, & Kaufman, 1995; Poulin-Dubois & Shultz, 1990; Rakison & Poulin-Dubois, 2001; Rochat, Morgan, & Carpenter, 1997; Sommerville & Woodward, 2005; Spelke, Phillips, & Woodward, 1995; Surian, Caldi, & Sperber, 2007).

However, it remains less clear how an understanding of the fundamental principles underlying biological processes emerges in development. The animate–inanimate distinction may be a precursor to understanding the domain of living things, but this distinction can be easily approximated perceptually according to which objects move on their own, which objects engage in goal-directed actions, and which objects show other perceivable patterns of biological behavior or motion unique to living kinds. To proceed from an animate–inanimate distinction to an understanding of the living–nonliving distinction, one must group together entities that vary drastically on the perceptual level both statically and dynamically (plants and animals against nonliving natural kinds and artifacts). This requires

---

This research was supported by the National Institutes of Health Grant R37-HD023922 to Frank C. Keil. We thank the staff, parents, and students at Bell-Graham, Gales Ferry, Guilford Lakes, and Thalberg elementary schools for their support. Special thanks to Yaw Anim, Amy Hart, and Jared Aguirre for their assistance with this research.

Correspondence concerning this article should be addressed to Jane E. Erickson, Department of Psychology, Yale University, 2 Hillhouse Ave., New Haven, CT 06520-8205. Electronic mail may be sent to jane.erickson@yale.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8101-0026

that one look past perceptual features to identify more fundamental characteristics of living things, such as physiological properties and processes they share in common, and the causal structures that explain these characteristics and behaviors (namely, that these entities are living and thus need to perform certain actions to maintain their survival in niches they typically inhabit).

One influential view claims that children do not have an understanding of biology and its underlying causal principles until late in childhood, and argues further that children first conceive of living things in a psychological, not biological, framework (Carey, 1985; Solomon, 2002; Solomon, Johnson, Zaitchik, & Carey, 1996). This view arises from the finding that children fail to group together plants with animals into one category (i.e., *living thing*) that guides their reasoning about biological parts and properties. Second, children reason about biological processes in terms of psychological states such as desires and beliefs. For example, when asked why various living things sleep or reproduce, children younger than 7 years old answer, ''Because they want to dream'' or ''Because they want to have babies,'' thereby making reference to desires and thus construing these processes as psychological, rather than biological, in nature. These sorts of studies seem to suggest that children do not have an understanding of biology—neither of what entities are alive nor an understanding of the causal mechanisms responsible for biological properties and processes. Instead, children's concept of *living thing* is akin to *animate thing* or an entity possessing mental states. All of the properties and behaviors of these entities are therefore supposedly understood to be the result of psychological, as opposed to biological, causal mechanisms.

Some researchers, however, argue that young children do possess a concept of *living thing* that includes both plants and animals while excluding nonliving things (Inagaki & Hatano, 2002; Keil, 1992, 1994; Waxman, Medin, & Ross, 2007). These researchers believe that children do have an understanding of some causal principles fundamental to the domain of biology, and furthermore, children understand that these principles differ from those in the physical and psychological domains.

Much of the debate revolves around induction projection tasks. In Carey's (1985) landmark series of studies young children projected biological properties based on psychological similarity to human beings, a pattern that was taken as showing that biology was not seen as distinct from psychology. Young children project a property, such as ''has a heart'' to dogs because they interpret hearts as supporting subjective well-being, something they see as a reasonable psychological state for a dog. They did not, however, seem to project ''has a heart'' to worms or other insects, a pattern that was explained on the grounds that such creatures are too psychologically dissimilar from humans. This pattern of projection, however, does not necessarily mean that children can only engage in psychological reasoning about biological properties. Projection experiments using a context/no-context manipulation find different results (Gutheil, Vera, & Keil, 1998; Inagaki & Hatano, 1996). For example, in one study's (Gutheil et al., 1998) biological context condition, children were told how the given property was biologically important to humans before being asked to project it onto other entities (e.g., ''This person eats because he needs food to live and grow. The food gives him energy to move. If he doesn't eat, he will die.''). In that context condition, both 4- and 5-year-olds properly extended this property to low-level organisms but not to nonliving things, unlike young children in Carey's studies or children in that study's no-context or social-context condition. These sorts of findings suggest that children do in fact have a concept of *living thing* as something separate from *animate* or *psychological thing*, but this concept needs to be made salient. In other words, psychology may be the default method of reasoning about the actions of most living things for young children, but a biological understanding may also be present and elicited in the right contexts.

The same message is emerging from cross-cultural studies, where certain cultures seem to provide contexts that early on make biological relations the default option in inductive reasoning tasks. Young children from more rural settings often show patterns of induction that seem to be based on biological principles (Medin & Waxman, 2007; Ross, Medin, Coley, & Atran, 2003). Indeed, it may be that urban and suburban contexts should be thought of as more atypical in that they reduce an emphasis on biological ways of construing the world, a pattern that converges with arguments that biological knowledge may have ''devolved'' in some developed areas of the West (Atran, Medin, & Ross, 2004).

More broadly, young children often seem to understand there are processes distinct to biological entities, and further, they see them as distinct from psychological ones. For example, children as young as 5 years seem to know that animals and plants, but not nonliving natural kinds and artifacts, grow in size over time (Inagaki & Hatano, 1996; Keil, 1983; Rosengren, Gelman, Kalish, & McCormick, 1991) and are capable of self-healing (Backscheider, Shatz, & Gelman, 1993; Inagaki, 2001; Inagaki & Hatano, 1996). Children also show different attribution patterns for biological and psychological processes across different entities (Coley, 1995; Jipson & Gelman, 2007), suggesting they do not treat these processes as one and the same.

In addition to studying which entities young children think are susceptible to biological processes, another way to discover whether or not they see a fundamental distinction between the biological and psychological domains is to look more closely at what children think are the causal mechanisms underlying biological phenomena. Are biological processes seen as being caused by mental states (psychology) or by physiological mechanisms (biology)? Preschoolers do seem to understand that parentage (biological factors) plays a greater role than friendship (social factors) in determining certain biological characteristics (Springer, 1992, 1996). Similarly 4- to 6-year-olds recognize that biological factors (e.g., eating and sleeping habits) play a greater role in one's susceptibility to illness than do social factors (e.g., lying; Inagaki, 1997). Preschoolers also rate explanations of contagions or contamination as better explanations than those of immanent justice for contracting an illness (Inagaki, 1997; Kalish, 1997; Siegel, 1988; Springer & Ruckel, 1992). Other studies have found that young children have a hard time conceiving of biological states (e.g., headaches, toe swelling) arising from psychological causes and that psychosomatic and psychogenic bodily reactions require a more sophisticated understanding seen only in adults (Notaro, Gelman, & Zimmerman, 2001; Shulz, Bonawitz, & Griffiths, 2007). Although in other cultural contexts young children may favor social and psychological explanations over biological ones (Astuti, Solomon, & Carey, 2004), it is clear that young children do have access to biological modes of explanation and they often can link these appropriately to the set of living things.

Because intention plays a large causal role in understanding psychological mechanisms, other studies have looked at whether or not children understand that mental processes are under voluntary control while bodily processes are not. A number of researchers (Inagaki, 1997; Inagaki & Hatano, 1993; Smith, 1978) have found that children 4- and 5-year-olds can correctly identify certain processes as voluntary (e.g., chewing) or involuntary (e.g., feeling pain). Similarly, other studies (Inagaki & Hatano, 1993; Lockhart, Chang, & Story, 2002; Lockhart, Nakashima, Inagaki, & Keil, 2008; Miller & Bartsch, 1997) have shown that children think physical characteristics are more difficult to change than psychological ones, in addition to each being modifiable through different means (mental effort for psychological characteristics and physical practice for physical characteristics). These results suggest that children are able to differentiate between bodily and mental processes, while also having some understanding that the latter are under psychological control whereas the former are not.

There are also specific indications that young children may be relying on causal intuitions when they contrast biological and psychological process. One series of studies has looked at children's understanding of trait origins and the ways in which different kinds of interventions might be connected to psychological and biological phenomena (Lockhart, Aw, & Essig, 2004; Lockhart, Nakashima, & Inagaki, 2004; Lockhart et al., 2008; Nakashima, Lockhart, & Inagaki, 2003). For example, Lockhart et al. (2004) showed that children as young as 5 years of age linked the origins of psychological traits more to effort and instruction and the origins of biological traits more to inborn factors. These beliefs about differential origin accounts were then mirrored by beliefs about interventions, one of the hallmark ways of assessing causal reasoning. When asked what sorts of interventions would be most effective in changing traits, children at all ages thought that taking medicine or simply waiting for maturation to take its course would be more effective in changing physical traits than psychological ones. In contrast, interventions involving effort or instruction were seen as more influential for changing psychological traits.

All of the above studies suggest that by the time children are in kindergarten, they understand that animals and plants are similar in comparison with nonliving things by virtue of having capacities such as reproduction, growth, self-healing, susceptibility to illness, and so on. However, the majority of these studies tended to either test only one biological phenomenon at a time or simply looked to see if children grouped plants with animals, as against nonliving things, when making inductive inferences. The extent to which children's understanding of the biological domain is systematic and coherent remains unclear. Do kindergarten-aged children see some degree of unification and clustering to biological processes, rather than reacting to isolated instances and comparing them on a case-by-case basis? To what extent do young children see these mechanisms and phenomena as hanging together in a larger system unified by causal principles that are different from those operating in other domains?

The following studies address these questions by giving children processes that span the entire domains of biology and psychology, and asking whether or not children classify these processes as belonging to fundamentally different categories distinguished by their underlying causal mechanisms. Experiments 1 and 2 present participants with triads of both human and animal biological and psychological processes and ask which two members of the triad belong together, allowing us to look at the clustering of processes within both domains. Experiments 3 and 4 further examine whether this clustering is based on simple associations or stems from a deeper understanding of the causal principles distinct to each domain. We test participants' understanding of these causal principles by having them reason about interventions, a method often used to assess knowledge of causal relations in various domains (Hagmayer, Sloman, Lagnado, & Waldmann, 2007; Kushnir & Gopnik, 2005; Kushnir, Gopnik, Schulz, & Danks, 2003; Sommerville, 2007; Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003; Woodward, 2003). Indeed, judgments about interventions have often been claimed to be at the heart of identifying underlying causal relations (Woodward, 2003).

Based on the research just reviewed we hypothesize that children will succeed in making a general domain distinction among an array of biological and psychological phenomena by the time they are in kindergarten, thus displaying an understanding of biology and psychology as two fundamentally separate, but internally coherent, domains with different underlying causal structures. We further hypothesize that children may display this competence without explicitly understanding the detailed mechanisms related to those properties. This second prediction follows from studies showing that even adults often have strikingly incomplete understandings of mechanisms in various domains (Rozenblit & Keil, 2002) and that children can cluster phenomena in broad domains such as physical mechanics and social behavior while being unable to explain these phenomena (Keil, Stein, Webb, Billings, & Rozenblit, 2008).

Experiment 1a

Method

Participants. Participants were 24 kindergartners (M = 5 years 9 months) and 24 undergraduates with approximately an equal number of males and females in each age group. The children were recruited from local elementary schools and were predominantly from a White upper-middle-class background. The parents of all children gave written consent and each child agreed to participate. The kindergartners were tested in a quiet room at their school and were compensated with a sticker and a certificate. The undergraduates were recruited from university sign-up sheets and received minimal compensation (a candy bar) for their participation.

Materials. We used the deference method (Danovitch & Keil, 2004; Keil et al., 2008; Lutz & Keil, 2002) to discover the extent to which kindergartners differentiate biological from psychological processes. This method has been successful in discovering how children understand knowledge clusters within different domains. By investigating to whom children defer about a domain-specific process we can gain a greater understanding of what comprises a coherent domain of knowledge for young children. For example, Lutz and Keil (2002) found that preschoolers appropriately deferred to either a mechanical or a biological expert when given phenomena within the two domains, suggesting they are able to differentiate between biological and physical processes. Additionally, Danovitch and Keil (2004) demonstrated that children come to prefer consulting experts with knowledge of underlying scientific principles, as opposed to experts of trivia knowledge, and Keil et al. (2008) showed that even kindergarteners are sensitive to underlying causal structures in many broad domains corresponding roughly to the natural and social sciences. These results suggest that children understand there are different domains of knowledge structured according to deep causal principles. Moreover, children expect that experts who have knowledge of those principles should also have an understanding of a very broad domain of phenomena arising from them. That is, even young children realize that knowledge of certain key causal relations in a domain provide a person with a generative capacity to understand a potentially unlimited range of domain-specific phenomena that are governed by those principles.

In the present study we used the deference task to assess participants' understanding of biology and psychology as two distinct domains. Participants were given 16 questions phrased in the following way: "If I want to know about X, should I ask someone who knows about Y or someone who knows about Z?" In half of the questions, Y was a biological process and Z was a psychological process; this order was reversed for the remaining half. For example, a biological question would look like the following: ''If I want to know why people need to drink lots of water (biological), should I ask someone who knows why people get stuffy noses (biological), or someone who knows why people are bullies (psychological)?'' A psychology question would be phrased in the same manner but with a psychological process being questioned: ''If I want to know why people sometimes think out loud (psychological), should I ask someone who knows why people need sleep to grow taller (biological), or someone who knows why people like to learn new things (psychological)?''

We chose 24 biological and 24 psychological processes we thought spanned the full range of both domains. These processes were further divided into two subcategories within each domain: *health* and *illness* in biology and *cognitive* and *emotional* in psychology. The ''health'' subcategory in biology included items such as growth, general nutrition, benefits of exercise, and so on, whereas the ''illness'' subcategory included processes dealing with aging, bodily malfunction, sickness, and self-healing. In the ''cognitive'' subcategory of psychology, we included things such as learning, memory, reasoning, and problem solving, whereas the ''emotion'' subcategory included processes that focused more on personality traits, social relationships, emotions and motivation. We realize there may be more than one way to split up the domains of biology and psychology, but we chose these particular subcategories because they are both salient in the lives of young children (there are many everyday instances of each) and because they are often clearly contrasting areas in adult discussions of these domains. Thus, there are clearly distinct adult groups of experts on wellness, pathology, social psychology, and cognitive psychology.

We included items from two subcategories within each domain not only to ensure we chose a wide range of processes but also to ensure we could compare performance on within- and between-subcategory questions. Half of the questions included in this task were *within subcategory*, meaning that participants need only match processes within the same subcategory. For example, a psychology question asking about a cognitive process would have two answer options: a biological process and a psychological process from the same ''cognitive'' subcategory. The other half of the questions was *between subcategory*, requiring that participants abstract across the full domain in order to answer the question correctly. For example, a psychological between-subcategory question that asked about a cognitive process would have two answer options: a biological process and a psychological process from a different psychological subcategory (emotional). In this case, participants have to recognize how two very different processes (cognitive and emotional processes) are fundamentally similar to one another, and different from those in the domain of biology.

The between-subcategory questions are the strongest test of our hypothesis, namely, that kindergartners recognize a common domain across these two distinct subdomains. However, we chose to include the within-subcategory questions because, it is possible that kindergartners are not yet able to recognize the similarities among all biological processes, or all psychological processes, but are able to recognize similarities on a more local level—within a particular subcategory. For example, we think young children will notice that learning and memory are similar to one another (both cognitive activities) before they are able to grasp that learning and memory are similar to other psychological processes such as feeling scared or wanting to help others. We could find that participants are able to recognize the biology–psychology distinction on a local level when presented with processes that are similar within each domain. In other words, it is possible that kindergartners recognize that cognitive processes are similar to one another, and different from biological ones, but fail to understand that all psychological processes are fundamentally similar in that they share a common set of underlying causal principles related to one's mental life that are different from those at the core of biological phenomena.

Eight of the processes in each domain (four from each subcategory) were used as the target in question whereas the remaining 16 processes were included as answer options. A complete list of the stimuli can be found in Appendix A with an asterisk next to those processes that were used as questions.

The answer choice pairings were counterbalanced across two conditions so one group of participants saw a particular pairing with a biological question whereas the other group of participants saw the same pairing but with a psychological question. The same questions and answer pairings were used with all participants, but in different combinations between the two counterbalanced groups. This counterbalanced design ensured that neither of the answer options was inherently preferable to the other.

We were careful in constructing the question–answer pairings to eliminate the possibility that participants could use a simple heuristic to cluster processes within each domain. For example, although we incorporated anatomical terms into the psychology items, they were more frequent among the biology items. However, if a body part was mentioned in the biology target question, it was not mentioned in the corresponding biology answer option and vice versa. Furthermore, all of the items were subject to latent semantic analysis to make sure neither of the answer options was more statistically related to the target question than the other. Given a large and representative corpus of text, latent semantic analysis looks at all of the contexts in which a given word is or is not used to determine where each word in the English language stands in relation to all other words in terms of how often such words co-occur either directly or indirectly through joint co-occurrence with other words (Deerwester, Dumais, Landauer, Furnas, & Harshman, 1990; Landauer, 1998; Landauer, Foltz, & Laham, 1998). This method therefore helps determine how close words are to each other in a multidimensional, correlational space. For our particular stimuli, this method was used to verify that the set of words in each answer option was not any more related to the set of words in the question given the many other contexts in which these sets of words have appeared (both answer options were equally correlated with the question). Such computations can be made at the following Web site: http://lsa.colorado.edu/. It might well be the case that in typical use, there would be higher levels of correlation between words used in a common domain and that such correlations might be a cue to relatedness. Here, however, we controlled for such statistical relatedness to ask if children could still cluster processes by using a deeper conceptual understanding that went beyond merely noting patterns of co-occurrence among lexical items.

Adult participants were given the questions in a paper-and-pencil task. The kindergartners were asked the questions by an experimenter, each question being accompanied by a pictorial representation of each process for memory purposes. The pictures were all acquired from Clip Art and controlled for gender, age, number of people in each picture, and any other surface features children might use to make a decision. See Experiment 1b for confirmation that kindergartners could not rely on the pictures alone to perform this task successfully.

Procedure. Before beginning the experiment, kindergartners were given a training task to make sure they understood the deference method being used. They were first asked if they ever had a question about how things worked or why certain things happened. If they did not come up with something on their own, they were prompted by the experimenter with a question such as: ''Have you ever wondered why you need to eat vegetables?'' Once there was a question established, the experimenter inquired about who the child asked or would ask if they wanted to know the answer. The children were then instructed that in the following task they were going to be given questions and their job was to give the experimenter advice about who to ask in order to find the answer. They were told there would be two people the experimenter could pick from, and they should think really hard about which person the experimenter should ask as the two people ''know about different kinds of things.'' After children were given these instructions, they were presented with two crucial practice items—both of which they had to pass to continue on with the experiment. The first practice item was the following: ''If I want to know why it sometimes rains outside, should I ask someone who knows about the weather or someone who knows about sports?'' The second practice item was more difficult because the answer options were both topically similar to the question, thus requiring that children use their knowledge of domain-specific principles to answer the question: ''If I want to know how a TV works, should I ask someone who knows all about the parts inside of a TV or someone who knows all about cartoons on TV?'' All of the kindergartners passed both practice items.

After completing the training items, the experimenter then proceeded to the test items reminding children there were no right or wrong answers. As each question was asked, a sheet with three pictures was placed in front of the child—one picture for the question and one picture for each answer option. After the child chose one of the experts, the pictures were removed, new pictures laid before the participant, and another question was asked. Throughout the experiment, participants were randomly asked to repeat the question and answer options, and sometimes to explain their answers. No feedback was given to the participants during this time. This questioning was simply to make sure that participants were engaged in the task and were able to remember the question and each answer option. Adult participants completed this task on paper in a quiet room in our laboratory. They were given written instructions directing them to circle the best answer out of the two given.

Results and Discussion

Accuracy was computed as the average proportion of questions for which participants correctly matched processes within the two domains (biology with biology and psychology with psychology). No differences were found between the two groups that counterbalanced the question and answer pairings for the kindergartners ($M = .77$, $SD = .01$ vs. $M = .82$, $SD = .02$), $t(22) = .92$, $p = .37$, or adults ($M = .96$, $SD = .002$ vs. $M = .94$, $SD = .004$), $t(22) = .94$, $p = .36$, so we collapsed the scores within each age group for the remaining analyses. The results from Experiment 1, including a breakdown by domain (biology and psychology) and by type of question (within or between subcategory), are depicted in Figure 1.

Adults were used as a baseline to compare the performance of kindergartners. Both adults ($M = .95$, $SD = .053$), $t(23) = 41.40$, $p < .001$, and kindergartners ($M = .80$, $SD = .124$), $t(23) = 11.70$, $p < .001$, performed significantly above chance on the overall task, although adults performed better than kindergartners, $t(46) = 5.64$, $p < .001$. Performance was also assessed separately for the biology and psychology items. Kindergartners and adults performed significantly above chance on both the biology—kindergartners: ($M = .83$, $SD = .13$), $t(23) = 12.09$, $p < .001$; adults: ($M = .97$, $SD = .07$), $t(23) = 33.78$, $p < .001$—and psychology items—kindergartners: ($M = .77$, $SD = .21$), $t(23) = 6.39$, $p < .001$; adults: ($M = .94$, $SD = .072$), $t(23) = 29.70$, $p < .001$—with no differences found between the two domains—kindergartners: paired-$t(23) = 1.15$, $p = .26$; adults: paired-$t(23) = 1.63$, $p = .12$. Furthermore, all scores within each of the subcategories were significantly above chance for both groups of participants.

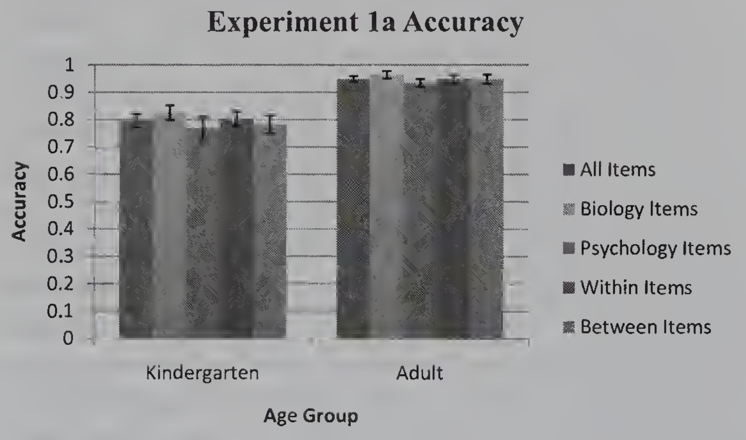


Figure 1. Experiment 1a accuracy scores for both kindergartners and adults. The overall accuracy is represented in the first bar for both age groups. The four remaining bars represent the accuracy broken down in two ways: by type of process (biological or psychological) and by type of question (within or between).

We further analyzed the within- and between-subcategory questions separately to see whether the children's success on the task was driven solely by their performance on within-subcategory questions. The kindergartners and adults performed significantly above chance on both the within- and between-subcategory items—kindergartners: ($M = .80$, $SD = .13$), $t(23) = 11.71$, $p < .001$, and ($M = .78$, $SD = .16$), $t(23) = 8.57$, $p < .001$, respectively; adults: ($M = .95$, $SD = .07$), $t(23) = 30.92$, $p < .001$, and ($M = .95$, $SD = .08$), $t(23) = 27.30$, $p < .001$, respectively—and performed equally well on both types of questions—kindergartners: paired-$t(23) = .68$, $p = .504$; adults: paired-$t(23) = .019$, $p = .99$.

These results suggest that kindergartners accurately distinguish biological from psychological processes. They performed equally well on the biology and psychology items, including consistent performance across each of the subcategories. Furthermore, their similar performance on the within- and between-subcategory items suggests that they have a robust understanding that all of the processes within each of these domains cluster with one another allowing the children to consistently classify them together.

Experiment 1b

The pictures used with the kindergartners in Experiment 1a were intended solely as memory aides and were chosen so as to offer no hints to the correct answer. However, we found while running the experiment that most of the kindergartners answered the questions by pointing at one of the two pictures representing the answer options. We decided to present a separate group of kindergarten participants with the same pictures, but this time without the verbal descriptions of the processes they represent. If kindergartners are finding superficial cues in the pictures to help them pair the correct answer option with the target in question, then they should perform equally well (and above chance) when presented with the pictures alone. If, however, the pictures are solely functioning as memory aides and do not give children any hints to the correct answers, then they should perform no better than chance (50% accuracy) on the task using only the pictures as their basis for judgment.

Method

Participants. Participants were 15 kindergartners (M = 5 years 9 months; 7 males, 8 females) who were recruited from local elementary schools and were predominantly from a White upper-middle-class background. The parents of all children gave written consent and each child agreed to participate. The kindergartners were tested in a quiet room at their school and were compensated with a sticker and a certificate.

Materials and procedure. In Experiment 1a, three pictures (representing three processes) were presented to kindergarten participants during each question. The first picture represented the target process in question, whereas the second two pictures represented the processes in each answer option. For example, three pictures accompanied the following item: "If I want to know why people need to drink lots of water, should I ask someone who knows why people get stuffy noses, or someone who knows why people are bullies?" The first picture depicted drinking lots of water, whereas the other two depicted stuffy noses and people being bullies.

These same picture triads were presented to participants in Experiment 1b but without the verbal descriptions of the processes they represent. These participants were instead asked which of two pictures "went with" or was "more similar to" the third target picture. A total of 32 picture triads (16 from each counterbalanced group in Experiment 1a) were presented to the participants.

Results

Accuracy was computed as the average proportion of questions for which participants matched pictures depicting processes within the same domain (biology with biology and psychology with psychology). Participants performed no better than chance on the triad task when only using the pictures as their basis for judgment ($M = .53$, $SD = .07$), $t(14) = 1.04$, $p = .13$. This remained true when the data were analyzed separately for the picture triads in each counterbalancing condition ($M = .51$, $SD = .106$), $t(14) = .34$, $p = .74$, and ($M = .55$, $SD = .08$), $t(14) = 2.15$, $p = .09$, respectively.

We can conclude from Experiment 1b that kindergartners could not have performed well in the first experiment if they only relied on the pictures to match various processes. This further demonstrates that kindergartner's success on the triad task in Experiment 1a results from a deeper understanding of biological and psychological processes as belonging to two distinct domains.

Discussion

Experiment 1 demonstrates that kindergartners appropriately cluster biological processes with one another while separating them from psychological ones. These results suggest that young children do sense the coherence among biological processes.

Although we chose a wide range of biological and psychological processes in Experiment 1, we were limited in that the biological referent was always a human being. In order to include the greatest range of processes in both domains, we expanded our items in Experiment 2 to also include a more diverse set of biological entities.

Experiment 2a

Method

Participants. Participants were 24 kindergartners (M = 5 years 10 months) and 24 undergraduates with approximately an equal number of males and females in each age group. The scores from two additional kindergartners were dropped, one for failing to complete the experiment and the other for failing the training task and always picking the second answer option for every question. The children were recruited from local elementary schools and were predominantly from a White upper-middle-class background. The parents of all children gave written consent and each child agreed to participate. The kindergartners were tested in a quiet room at their school and were compensated with a sticker and a certificate. The undergraduates were recruited from university sign-up sheets and received minimal compensation (a candy bar) for their participation.

Materials. We used the same deference task as in Experiment 1 but with two major changes. First, instead of focusing on human biology and psychology, we expanded our items to include a range of animals and, thus, a greater range of processes within each domain. Second, we eliminated the within-subcategory items and only included questions that were between subcategory, as these are the strongest test of our hypothesis.

The animals were chosen based on three criteria: (a) they had to plausibly possess psychological states, (b) they had to represent a range of animals to ensure a greater range of biological and psychological processes, and (c) young children should be at least somewhat familiar with them (even if never having a direct encounter). The following were the eight animals we chose: dog, cat, horse, monkey, elephant, camel, fish, and bird.

We once again divided the various biological and psychological processes into two subcategories per domain. The psychological subcategories were the same as in Experiment 1: *cognitive* and *emotional*. The biological subcategories, however, were changed. One subcategory, *anatomy*, focused on an animal's anatomical features and their adaptive value. The second subcategory, *physiology*, consisted of physiological processes such as growth, digestion, breathing, and body temperature regulation. We chose to divide the biological processes into these two subcategories because one concerns static structures and the other concerns dynamic processes, presumably one of the largest possible contrasts within the domain of biology.

There were a total of 24 biological and 24 psychological processes (12 in each subcategory) in Experiment 2. Each animal was associated with 6 processes—3 biological and 3 psychological. This allowed us to present participants with both a biological and psychological question for each animal (the other 4 processes were used as answer options for these two questions). This means that the same animal was mentioned in both the question and its two answer options. For example, if we presented participants with the question, "If I want to know how cats remember their way home (cat–cognitive), who should I ask?" both answer options would also apply to cats: "Someone who knows why cats have blood in their bodies (cat–anatomy), or someone who knows why cats have fun playing with yarn (cat–emotional)?" Keeping the animal consistent between the question and its answer options ensured that participants were choosing their answer based on the similarities and differences between the *processes* presented rather than the animals subject to those processes.

Each participant was given a total of 16 questions: 8 biological and 8 psychological (4 from each subcategory). A complete list of the stimuli can be found in Appendix B with an asterisk next to those processes that were used as questions.

Just as in Experiment 1, latent semantic analysis was used to make sure that neither of the answer options was more strongly associated with the process in question. We also included two counterbalancing conditions as in Experiment 1 (one group of participants would see a pair of answer options with a biology question whereas the other group of participants would see the same pairing with a psychology question) to make sure there was no inherent preference for either of the answer options. As in Experiment 1, we also controlled for other factors children could use as simple heuristics in making their judgments.

Adult participants were once again given the questions in a paper-and-pencil format. The kindergartners were asked the questions by the experimenter, each question accompanied by a pictorial representation of each process for memory purposes.

Procedure. The procedure was identical to that of Experiment 1 for both the children and adults. This included using the same training task and two crucial practice items with the children.

Results

As in Experiment 1, accuracy was computed as the average proportion of questions for which participants correctly matched processes within the two domains. No differences were found between the two different groups that counterbalanced the question and answer pairings for the kindergartners ($M = .75$, $SD = .07$ vs. $M = .73$, $SD = .10$), $t(22) = .53$, $p = .60$, or adults ($M = .82$, $SD = .07$ vs. $M = .87$, $SD = .10$), $t(22) = 1.23$, $p = .23$, so we collapsed the scores within each age group for the remaining analyses. The results from Experiment 2 are depicted in Figure 2.

Once again, adults were used as a baseline to compare the performance of kindergartners. Both adults ($M = .85$, $SD = .09$), $t(23) = 19.33$, $p < .001$), and kindergartners ($M = .74$, $SD = .09$),

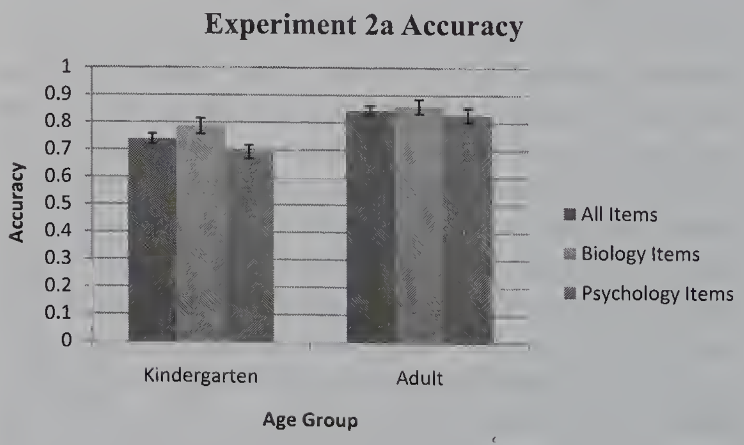


Figure 2. Experiment 2a accuracy scores for both kindergartners and adults. The overall accuracy is represented in the first bar for both age groups. The second two bars represent accuracy for biology and psychology items separately.

$t(23) = 13.65$, $p < .001$, performed significantly above chance on the overall task, although once again adults performed better than the kindergartners, $t(46) = 4.13$, $p < .001$. We also examined performance on the biology and psychology items separately. Adults scored significantly above chance on both the biology ($M = .86$, $SD = .13$), $t(23) = 10.12$, $p < .001$, and the psychology questions ($M = .83$, $SD = .12$), $t(23) = 12.73$, $p < .001$, and performed equally well in both domains, paired-$t(23) = .82$, $p = .42$. Kindergartners scored slightly better on the biology than on the psychology items ($M = .79$, $SD = .14$ and $M = .70$, $SD = .12$, respectively) paired-$t(23) = 2.33$, $p = .03$; however, they performed significantly above chance in both domains—biology, $t(23) = 10.12$, $p < .001$, and psychology, $t(23) = 7.83$, $p < .001$. Furthermore, both kindergarten and adult accuracy reached significance for all four biological and psychological subcategories (anatomy, physiology, emotion, and cognition).

These results demonstrate that kindergartners are capable of distinguishing biological from psychological processes, even when presented with a wide range of living organisms. Furthermore, all of the questions used in Experiment 2a were between subcategory, requiring that participants abstract across the full range of processes within each domain.

Discussion

Experiment 2a replicated the results from Experiment 1a, demonstrating that kindergartners are capable of making very subtle distinctions between biological and psychological processes, distinctions they would only be capable of making if they sensed some underlying coherence uniting all of the biological processes into a single domain.

Experiment 2b

Just as Experiment 1b ensured that participants could not rely solely on the pictures to perform successfully in Experiment 1, we included the same test for the pictures used in Experiment 2.

Method

Participants. Participants were 15 kindergartners (M = 5 years 10 months; 8 males, 7 females) who were recruited from local elementary schools and were predominantly from a White upper-middle-class background. The parents of all children gave written consent and each child agreed to participate. The kindergartners were tested in a quiet room at their school and were compensated with a sticker and a certificate.

Materials and procedure. The same picture triads from Experiment 2a were presented to participants without the verbal descriptions of the processes they represent. These participants were asked which of two pictures "went with" or was "more similar to" the third target picture. A total of 32 picture triads (16 from each counterbalanced group in Experiment 2a) were presented to the participants.

Results

As in Experiment 1b, accuracy was computed as the average proportion of questions for which participants matched pictures depicting processes within the same domain. Once again, the kindergartners performed only at chance when matching pictures in the triads without knowing the processes they represent ($M = .50$, $SD = .07$), $t(13) = .23$, $p = .82$, even when analyzed separately for both counterbalanced groups ($M = .55$, $SD = .12$), $t(13) = 1.65$, $p = .12$, and ($M = .46$, $SD = .13$), $t(13) = 1.26$, $p = .23$, respectively.

These results confirm our hypothesis that the pictures used in Experiment 2a were not driving the kindergartner's successful performance on the task. Rather, it was knowledge of the processes these pictures represent that allowed children to appropriately cluster items into the domains of biology and psychology.

Discussion

Both Experiments 1 and 2 demonstrate that kindergartners appropriately cluster biological processes with one another while separating them from psychological ones. These results suggest that young children sense some deeper similarity among processes in the biological domain. In other words, even kindergartners understand more than the simple idea that illness is caused by contagions or that physical traits are determined by parentage not friendship. These children are capable of looking at a diverse range of biological processes and classifying them together, separate from processes in the domain of psychology.

However, two critical questions remain to be answered before we can conclude that children recognize the fundamentally different nature of biological and psychological phenomena. First, *what* allows children to distinguish among them? Do children have a clear understanding of how they differ or do they simply sense that one biological process is *more* similar to another process within biology than one that is psychological? In other words, is this a principled difference or simply one that is a matter of degree? Second, *what* do children think are the underlying similarities uniting the range of processes within each domain? We have yet to show that kindergartners understand that phenomena within these two domains are *fundamentally* similar—stemming from similar causal mechanisms. Experiments 3 and 4 address this issue.

Experiment 3

Method

Participants. Participants were 16 kindergartners (*M* = 5 years 10 months) and 16 undergraduates with approximately an equal number of males and females in each age group. The scores from 2 additional kindergartners were dropped for failure to complete the experiment. The children were recruited from local elementary schools and were primarily from an upper-middle-class background. The parents of all children gave written consent and each child agreed to participate. The children were tested in a quiet room at their school and were compensated with a sticker and a certificate. The undergraduates were recruited from university sign-up sheets and received minimal compensation (a candy bar) for their participation.

Materials. In Experiment 3, we used many of the same processes from Experiment 1 but presented them to participants in a different context. In the two previous experiments participants were asked to match one of the answer options with the target in question. In this experiment, however, participants were given two processes in the question and asked how *causally* similar they were to a third. Both processes in the question were from the same domain (both biological or both psychological), whereas the third process was either from a different subcategory in that domain or from a different domain altogether. The questions asked in Experiment 3 had the following format: "Something happened to Mary a long time ago. Ever since then, Mary is not as lazy as she used to be (psychological). And ever since then, Mary helps others more than she used to (psychological). How much do you think what happened to Mary also made a difference for how often she gets sore throats (biological)?"

Participants were asked to give a rating on a scale of 1–5 (1 meaning there would be no effect on the third item and 5 meaning there would be a very large effect on the third item). For half of the items, two psychology processes were stated in the question whereas for the other half, the two processes were biological. We used two, rather than one, items in each question to demonstrate that the source of change to one process can result in a change to another. This was intended to demonstrate the salience and plausibility of causal influence. Also, for half of the psychology questions, the target phenomenon to be rated was biological (between domain) whereas in the other half it was psychological (within domain). The biology questions had the same structure. Just as in Experiments 1 and 2, we performed latent semantic analysis on all of the items to make sure the target process was equally related to the processes in question for both the within- and between-domain items. There were a total of 12 questions in this first task.

These questions were designed to tap into children's understanding of the causal mechanisms behind biological and psychological processes. Children should rate the target item as having a high likelihood of being affected if it corresponds to the same domain of processes (biology or psychology) presented in the question and low if it does not. This is because if some change occurs that affects the causal mechanism in a given domain, many processes within that domain (and not other domains) are likely to be affected.

As mentioned in the Introduction, one crucial way to assess whether children are invoking causal reasoning when thinking about biological and psychological processes is to ask them to reason about interventions. To that end, in the second part of Experiment 3, we used a method developed by Lockhart and her colleagues (Lockhart et al., 2002; Lockhart et al., 2008; Lockhart, Aw, et al., 2004; Lockhart, Nakashima, et al., 2004). These researchers conducted the first studies showing how an intervention orientation results in young children clearly distinguishing psychological from biological phenomena. In one set of studies Lockhart, Nakashima, et al. (2004) showed that children as young as 5 years of age saw biological properties (e.g., height and vision) being more influenced by biological interventions than were psychological properties. Similarly, they found that young children saw psychological properties (e.g., being shy or intelligent) as being more influenced by psychological interventions. In addition, they showed that causal reasoning seems to be explicitly elicited in this kind of task format as evidenced by spontaneous comments made by children that invoked causal reasons. This finding on interventions converged strongly with another finding in Lockhart et al. (2004) in which young children attributed the origins of psychological trait differences more to effort and instruction, and the origins of biological trait differences more to inborn factors and a lack of maturation. We used a method similar to that of Lockhart and her colleagues to determine whether the knowledge clustering found in Experiments 1 and 2 was indeed making reference to causal principles.

The second part of this experiment, therefore, explicitly questioned participants about the causal mechanisms underlying the domains of biology and psychology. Participants were given all 12 descriptions from the first part of the experiment (e.g., "Something happened to Mary a long time ago. Ever since then, Mary is not as lazy as she used to be. And ever since then, Mary helps others more than she used to."), but this time participants were asked: "What do you think happened to Mary? Do you think she took a special pill or a special class?" Participants were asked to choose which of the two sources of influence was more likely to apply to the scenario in question. Through questioning participants about the source of influence, we were able to see whether they consistently identify two different types of causation (physical or mental) whenever biological or psychological processes are manipulated.

Procedure. Before beginning the experiment, kindergartners were given a training task to make sure they understood both the questions being asked and the rating scale being used. They were first given a description about how certain changes will or will not influence other things about a particular entity: "So what we're going to do today is talk about things that change. When we make a change to something, sometimes other things also change about it, but not always. For example, if we sharpen a pencil then it will also get shorter and it will also write better. But if we just put a sticker on the pencil, it won't get any shorter or write better. Nothing else changes about the pencil other than the sticker being on it." They were then given two scenarios to practice using the 5-point scale. In one scenario the experimenter asked how likely a change in engine and a change in tires would influence how fast a car could go. In the second scenario the experimenter asked how likely a change in color and a change in seat comfort would influence how fast a car could go. After the participant answered these questions, whether correctly or incorrectly, the experimenter explained why the first scenario would likely make a big difference for the speed of a car while only a small difference or no difference would be seen in the second scenario. These two scenarios helped to demonstrate the causal process behind a change in some entity.

The last practice item given to participants required they not only understand the causal process of change but also be able to infer such a process by its results, given this was the format of the questions in the rest of the experiment. Participants were told that something happened to Carrie a long time ago. "Ever since then Carrie owns a better car than she used to and ever since then Carrie has a nicer house than she used to. How much do you think what happened to Carrie also made a difference for how big of a TV Carrie has?" This question is different from the former scenarios because in this one, owning a nicer house and owning a better car does not *directly* cause Carrie to own a better TV. Therefore participants must infer a third causal mechanism that can influence both of the items given, and then determine whether it would also influence the third target item. The entire training procedure took no more than 5 min, and children's ease with these training questions provides additional support that they understand and reason about underlying causal principles.

Throughout the experiment, participants were randomly asked to repeat the question and answer options and sometimes to explain their answers. No feedback was given to the participants during this time. This questioning was simply to make sure that participants were engaged in the task and were able to remember the question and each answer option.

Adult participants completed this task on paper in a quiet room in our laboratory. They were given written instructions directing them to circle a number on the 5-point scale in the first task and one of the two answer options (special pill or special class) in the second task. The kindergartners were asked the questions by the experimenter, each question being accompanied by a pictorial representation of each process for memory purposes (the same pictures from Experiment 1).

Results

Results from Experiment 3 can be seen in Figures 3 and 4. For the first task in this experiment, the data were analyzed according to whether the item was a within- or between-domain

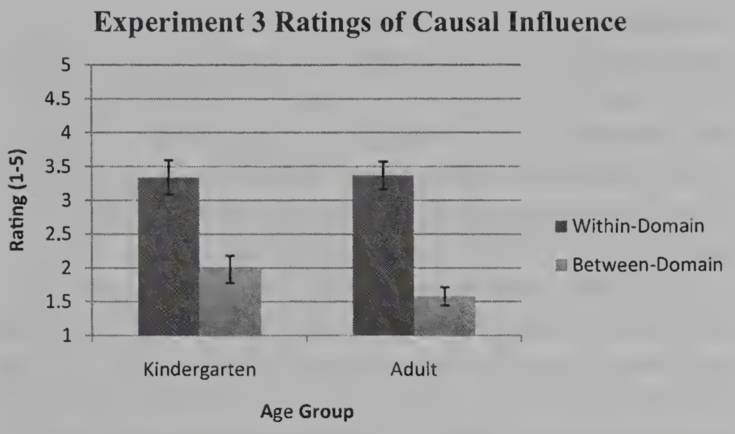


Figure 3. Experiment 3 ratings of causal influence for both within- and between-domain questions. Participants rated on a scale of 1–5 the extent to which changes in processes in one domain affected another process either in the same (within) or in a different (between) domain.

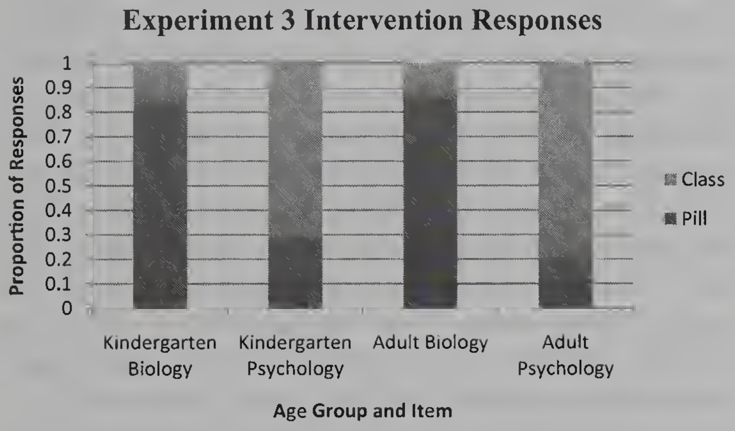


Figure 4. Results for the second part of Experiment 3. Each bar represents the kindergarten and adult responses according to the type of process in question: biological or psychological. Participants chose from two options—taking a "special class" or taking a "special pill"—as the most influential in changing that particular biological or psychological process.

question. The within-domain questions were those in which the target process to be rated was within the same domain as the processes presented in the question. The between-domain questions were those in which the target process to be rated was from a different domain than those previously presented. We expected that the within-domain questions would receive a higher rating on the 5-point scale than the between-domain questions given that if the causal mechanism is affected in one domain, it is likely to affect other processes within, but not outside, that domain.

Our predictions were supported for both adults and kindergartners. Adults gave an average rating of 3.37 out of 5 (SD = .82) for within-domain items and 1.58 (SD = .54) for between-domain items, paired-$t(15) = 9.83$, $p < .001$. Similarly, kindergartners gave higher ratings for the within items (3.34 out of 5, SD = 1.02) than the between items (M = 1.98, SD = .81), paired-$t(15) = 5.59$, $p < .001$. These differences were not driven by higher accuracy for either the biology or psychology questions. Both adults and kindergartners (M = 2.53, SD = .66 and M = 2.42, SD = .55 respectively), $t(15) = 1.49$, p = .16, performed equally well on the items within these two domains (M = 2.67, SD = .83 and M = 2.65, respectively), SD = .75, $t(15) = .29$, p = .78.

The second task required that participants choose the source of change for processes within both the biology and psychology domains. When a participant chose the biological source of change (took a special pill), his or her answer was coded as 0, while choosing the psychological source of change (took a special class) resulted in a score of 1. This allowed us to determine how often participants chose each answer for the biology and psychology items separately. Both adults and kindergartners predominantly chose the biological source of change for the biological processes (M = .146, SD = .120), $t(15) = 11.83$, $p < .001$, and (M = .167, SD = .202), $t(15) = 6.61$, $p < .001$, respectively, and the psychological source of change for the psychological processes (M = .792, SD = .177), $t(15) = 6.58$, $p < .001$, and (M = .708, SD = .177), $t(15) = 4.70$, $p < .001$, respectively.

Results from Experiment 3 reveal that kindergartners grasp the deeper causal mechanisms that structure the domains of biology and psychology. In the first task, we examined the extent to which participants understood that there are deep causal mechanisms affecting all of the processes within, but not outside, a given domain. Ratings of causal influence given by both kindergartners and adults demonstrate the depth of their understanding, giving higher ratings of causal influence to within- than between-domain processes. The second task asked participants to choose between two very different causal interventions (pill vs. class) as the source of change to various biological and psychological processes. Both kindergartners and adults more consistently identified the pill as the source of change for biological processes and the class as having a greater influence on psychological processes.

Discussion

The results from Experiment 3 indicate that even kindergartners have some sense of the fundamentally different causal mechanisms underlying biological and psychological processes, including the kinds of influence these mechanisms can have on such processes.

Just as with Experiment 1, we incorporated a large range of biological and psychological processes but were limited in that the referent was always a human being. We find it unlikely, but possible, that children were using a simpler heuristic on the last task of Experiment 3. It could be that children were basing their answers not on deep intuitions about the causal mechanisms in the biological and psychological domain but rather on social roles associated with the particular processes in question. Children may associate "taking a special pill" with doctors and "taking a special class" with teachers, thereby judging the processes with reference to what a doctor or teacher would know. In other words, children may associate taking pills with "catching colds" and both taking pills and catching colds with "things a doctor would know" without having any knowledge of the causal link between these items (that a doctor understands the causal mechanisms underlying illness in the body). It seems unlikely that children would associate biological processes like "sweating" with doctors or think that only psychological (not biological) processes would fall under the umbrella of knowledge possessed by teachers. However, we decided to directly address this issue in Experiment 4, using a wider range of biological and psychological processes across a wider range of biological referents so as to fully rule out this alternative interpretation of our results.

Experiment 4

Method

Participants. Participants were 16 kindergartners (M = 5 years 10 months) and 16 undergraduates with approximately an equal number of males and females in each age group. The children were recruited from local elementary schools and were predominantly from a White upper-middle-class background. The parents of all children gave written consent and each child agreed to participate. The kindergartners were tested in a quiet room at their school and were compensated with a sticker and a certificate. The undergraduates were recruited from university sign-up sheets and received minimal compensation (a candy bar or Snapple) for their participation.

Materials. The materials used in Experiment 4 were a subset of the animal biological and psychological processes from Experiment 2. We selected 16 processes (2 biological and 2 psychological) from four different animals (cat, bird, horse, and monkey). Four of the possible 6 processes were chosen from four of the possible eight animals. We chose those processes for which a causal intervention would make the most sense. For example, a process such as "Fish first grow in eggs" was not selected as it would be difficult to imagine how that process could be enhanced through any causal intervention, biological or otherwise. A process such as, "A cat grows sharp claws on its paws," however, could conceivably be influenced by some biological intervention—one that could potentially make claws grow quicker or sharper.

Participants were told that some change occurred for each of these processes and were questioned about the source of that change (the change presented always enhanced the process in question). For example, one of the cat biological processes from Experiment 2 was: "Cats grow sharp claws on their paws." In the present task the item was framed in the following way: "Something happened to this cat a long time ago, and ever since then, the cat grows even sharper claws on its paws. What do you think happened to the cat?" For all of the items, participants were presented with a forced choice between two answer options: the animal had an "operation" or the animal had "training." These two answer options mirrored the "pill" and "class" options from Experiment 3. This contrast was for the same purpose: to see whether participants consistently identify two different types of causation (physical or mental) whenever biological or psychological processes are affected. We changed the options from "pill" and "class" to "operation" and "training," however, as the processes now under consideration apply to animals, not human beings. We therefore needed more plausible means of influencing an animal's biology and psychology rather than having birds take pills and having horses sit through classes. We also thought that kindergartners, likely being unfamiliar with operations and even animal training, would not have strong associations between these answer options and either specific social roles or the particular biological and psychological processes presented throughout the experiment. In other words, we doubt that any kindergartner has previously heard of a monkey that underwent an operation to make its tail grow even longer. Having no preestablished associations to base their answers upon, participants must rely on a more abstract understanding of causal principles at the foundation of their biological and psychological knowledge to perform successfully on this task.

Participants were presented with all 16 processes and asked to choose between the two sources of change: operation or training. Adult participants were once again given the questions in a paper-and-pencil format. The kindergartners were asked the questions by the experimenter, each question being accompanied by a pictorial representation of each process (the same pictures from Experiment 2).

Procedure. The procedure used in Experiment 4 was nearly identical to the second task in Experiment 3. Participants were presented with 16 biological and psychological processes that had undergone some change. They were instructed to choose which of two causal mechanisms (operation or training) was the most plausible intervention for each of these items.

Kindergartners, being familiar with both pills and classes, required no explanation of these answer options in Experiment 3. However, the level of knowledge kindergartners have regarding operations and/or training is likely to be very little. To make sure all the kindergartners possessed a sufficient amount of knowledge about these answer options, we introduced them to the task in the following way: "Sometimes things happen and it makes animals different in some way. For example, something could happen to make an animal a faster runner. Or something could happen and it might make the animal nicer to other animals. I am going to tell you about different animals that have changed in some way. I'll tell you how they are different now than they used to be. I want you to tell me what you think happened to the animal to make it different, what you think made it change. There are two things that could have happened to the animal: it could have had an operation, or it could have had training. An operation is when people perform surgery on an animal; they go into the animal and change things deep inside of it. Training is when an animal learns new things from either people or other animals. So, I will tell you about an animal and something that changed about the animal. You will tell me if you think the animal is different because it had an operation, or because it had training. Pick whichever one you think made the animal different now than it used to be." These instructions were intended to give kindergartners enough information about operations and training that they could complete this task as they did previously with the "pill" versus "class" options. However, we left the instructions vague enough that kindergartners were not told explicitly *how* operations or training affected an animal's body or mind. It was up to the individual kindergartners to decide what would most likely affect any given process (e.g., such as getting scared): things changed deep inside the animal or things learned by the animal.

Adults received a shortened version of these instructions and performed the task with paper and pencil in a quiet room in our laboratory. The kindergartners were asked each question by the experimenter, and were tested in a quiet room at their elementary school.

Results

Results for Experiment 4 can be seen in Figure 5. Participants were asked to make a forced choice about the source of change to biological and psychological processes. When a participant chose the biological source of change (had an operation), his or her answer was coded as 0, while choosing the psychological source of change (had training) resulted in a score of 1. This allowed us to determine how often participants chose each answer for the biology and psychology items separately. Both adults and kindergartners predominantly chose the biological source of change for the biological processes ($M = .15$, $SD = .24$), $t(15) = 5.84$, $p < .001$, and ($M = .159$, $SD = .13$), $t(15) = 10.86$, $p < .001$, respectively, and the psychological source of change for the psychological processes ($M = .72$, $SD = .25$), $t(15) = 3.55$, $p < .003$, and ($M = .91$, $SD = .13$), $t(15) = 12.43$, $p < .001$, respectively.

These results demonstrate that even kindergartners understand the fundamentally different causal mechanisms influencing biological and psychological

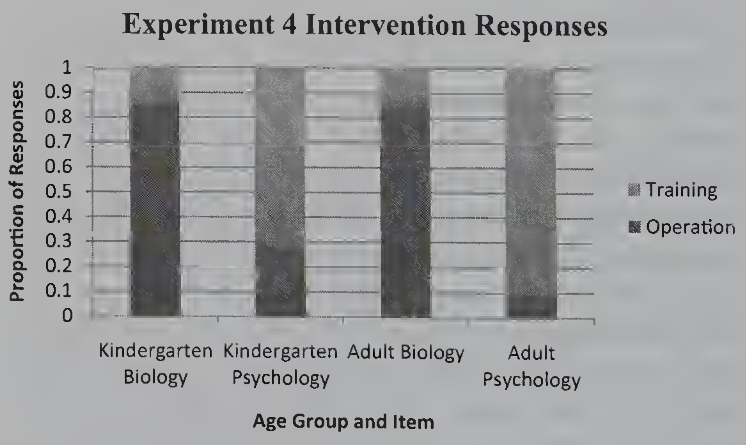


Figure 5. Results for Experiment 4. Each bar represents the kindergarten and adult responses according to the type of process in question: biological or psychological. Participants chose from two options—undergoing an "operation" or "training"—as the most influential on that particular biological or psychological process.

processes, even when taken outside the familiar realm of human biology and psychology and applied across a range of animals.

Discussion

Experiments 3 and 4 went beyond both Experiments 1 and 2, allowing us to gain insight into how children were clustering biological and psychological phenomena. Not only do children understand that these processes are different from one another (as was demonstrated in Experiments 1 and 2), but they understand these processes are *fundamentally* different, that is, based on very different causal mechanisms. Having this understanding of domain-specific causal mechanisms allows children to not only differentiate within- from between-domain processes but also to grasp the fundamental similarity uniting different processes within a given domain. In other words, children not only recognize that biology and psychology are different, but how biological processes are *similar* to one another, as are psychological ones. This understanding provides children with the means to reason successfully about how different processes within a domain are related to one another, how changes in one can lead to changes in another, and how these changes are mediated through specific causal mechanisms.

General Discussion

In addition to having knowledge within the domains of biology and psychology, as was illustrated in the introduction, it appears that children as young as 5 years old also have an understanding that the processes within these two domains are fundamentally similar in that they are subject to the same underlying causal principles. Children know more than simply that some things grow or reproduce, and thus they possess some biological knowledge that is independent of any psychological-explanatory framework. These four experiments suggest that children also understand that processes such as growth and reproduction are phenomena that should be integrated into one domain (biology) united by common causal mechanisms, ones that are different from those operating in the domain of psychology.

We can conclude from Experiments 1 and 2 that young children have strong intuitions about the contrast between the very large domains of psychology and biology. Furthermore, Experiments 3 and 4 demonstrate that children understand this contrast stems from different causal frameworks. What remains to be discovered is *what* children think are the underlying causal mechanisms and *how* they influence various biological processes. Do children have a rich understanding of these causal mechanisms and how they function, or do they only have a skeletal framework that allows them to make limited inferences and inductive generalizations? We have seen a theme emerge in the recent literature that young children often pick up on abstract patterns that differentiate categories before grasping all the concrete mechanisms. For example, Simons and Keil (1995) found that young children understand that living things, nonliving natural kinds, and artifacts have different insides from one another but were unclear about what exact insides are in each. The ability to identify broad domains through sensing abstract causal patterns may be critical to guiding children to look for more mechanistic contrasts as their knowledge develops further. Children may need to have some sense that broad domains of regularities have critically different causal bases before investing the time into trying to master the details of such differences. It may also be that certain kinds of general causal schemata, such as a teleological orientation, increases sensitivity to the biological–psychological contrast by highlighting broad causal-functional patterns that fit differently with two domains although not appealing to specific mechanisms (Opfer & Siegler, 2004).

There is further evidence that learning specific causal mechanisms may vary across cultures. These four studies were performed in the United States, but Lockhart et al. (2004) examined children's beliefs about biological and psychological interventions in both the United States and Japan (see also Lockhart et al., 2008). Although both cultural groups at all ages distinguished biological from psychological properties in terms of which interventions were seen as most important, there was also a cultural difference. Japanese participants were more likely than U.S. participants to select effort as the most effective intervention for changing physical traits whereas U.S. participants chose taking medicine more often than Japanese participants. This cultural difference raises the important question of whether the kinds of expertise clusters found here might be present in many cultures, but might stress different underlying causal influences as most central. Thus, young children may sense a key difference between broad domains but local cultures may guide them toward weighing different kinds of causal patterns as most essential.

Another remaining question concerns how young children come to understand the fundamental differences between these two domains in the first place. The particular tasks used in these studies were verbally laden and thus were not successful in pilot studies with preschoolers. However, future studies might reduce the verbal loads of such tasks to discover how children begin to make the distinction between processes in the biological and psychological domains. There is now considerable controversy about the idea that knowledge of biology stems from a previous understanding of psychology (Waxman et al., 2007). Studies with younger children may help determine the origins of this biological understanding, especially how children learn the causal mechanisms foundational for all biological processes and how these relate to the contrast between the psychological and biological world.

References

Astuti, R., Solomon, G. A., & Carey, S. (2004). Constraints on conceptual development: A case study of the acquisition of folkbiological and folksociological knowledge in Madagascar. *Monographs of the Society for Research in Child Development, 69*(Serial No. 277).

Atran, S., Medin, D., & Ross, N. (2004). Evolution and devolution of knowledge: A tale of two biologies. *Royal Anthropological Institute, 10*, 395–420.

Backscheider, A. G., Shatz, M., & Gelman, S. A. (1993). Preschoolers' ability to distinguish living kinds as a function of regrowth. *Child Development, 64*, 1242–1257.

Baillargeon, R. (1994). Infants' physical world. *Current Directions in Psychological Science, 13*, 89–94.

Bertenthal, B. I. (1993). Perception of biomechanical motions by infants. In C. Granrud (Ed.), *Visual perception and cognition in infancy* (pp. 175–214). Hillsdale, NJ: Erlbaum.

Bullock, M. (1985). Animism in childhood thinking: A new look at an old question. *Developmental Psychology, 21*, 217–225.

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.

Coley, J. D. (1995). Emerging differentiation of folkbiology and folkpsychology: Attributions of biological and psychological properties to living things. *Child Development, 66*, 1856–1874.

Danovitch, J. H., & Keil, F. C. (2004). Should you ask a fisherman or a biologist? Developmental shifts in ways of clustering knowledge. *Child Development, 75*, 918–931.

Deerwester, S., Dumais, S. T., Landauer, T. K., Furnas, G. W., & Harshman, R. A. (1990). Indexing by latent semantic analysis. *Journal of the Society for Information Science, 41*, 391–407.

Gelman, R., Durgin, F., & Kaufman, L. (1995). Distinguishing between animates and inanimates: Not by motion alone. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 150–184). Oxford, UK: Clarendon.

Gutheil, G., Vera, A., & Keil, F. C. (1998). Do houseflies think? Patterns of induction and biological beliefs in development. *Cognition, 66*, 33–49.

Hagmayer, Y., Sloman, S., Lagnado, D., & Waldmann, M. R. (2007). Causal reasoning through intervention. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, and computation* (pp. 86–100). Oxford, UK: Oxford University Press.

Inagaki, K. (1997). Emerging distinctions between naïve biology and naïve psychology. In H. M. Wellman & K. Inagaki (Eds.), *The emergence of core domains of thought: Children's reasoning about physical, psychological, and biological phenomena* (Vol. 75, pp. 27–44). San Francisco: Jossey-Bass.

Inagaki, K. (2001). *Young children's distinction between living kinds and complex artifacts: How living and nonliving entities recover from trouble.* Paper presented at the annual meeting of the Society for Research in Child Development, Minneapolis, MN.

Inagaki, K., & Hatano, G. (1993). Young children's understanding of the mind-body distinction. *Child Development, 64*, 1534–1549.

Inagaki, K., & Hatano, G. (1996). Young children's recognition of commonalities between animals and plants. *Child Development, 67*, 2823–2840.

Inagaki, K., & Hatano, G. (2002). *Young children's naïve thinking about the biological world*. New York: Psychology Press.

Jipson, J. L., & Gelman, S. A. (2007). Robots and rodents: Children's inferences about living and nonliving kinds. *Child Development, 78*, 1675–1688.

Kalish, C. (1997). Preschoolers' understanding of mental and bodily reactions to contamination: What you don't know can hurt you, but cannot sadden you. *Developmental Psychology, 33*, 79–91.

Keil, F. C. (1983). On the emergence of semantic and conceptual distinctions. *Journal of Experimental Psychology: General, 112*, 357–385.

Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.

Keil, F. C. (1992). The origins of an autonomous biology. In M. R. Gunnar & M. Maratsos (Eds.), *Modularity and constraints in language and cognition. The Minnesota symposia on Child Psychology* (Vol. 25, pp. 103–137). Hillsdale, NJ: Erlbaum.

Keil, F. C. (1994). The birth and nurturance of concepts by domain: The origins of concepts of living things. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture.* (pp. 234–254). Cambridge, UK: Cambridge University Press.

Keil, F. C., Stein, C., Webb, L., Billings, T., & Rozenblit, L. (2008). Discerning the division of cognitive labor: An emerging understanding of how knowledge is clustered in other minds, *Cognitive Science*, *32*, 259–300.

Kushnir, T., & Gopnik, A. (2005). Young children infer causal strength from probabilities and interventions. *Psychological Science*, *16*, 678–683.

Kushnir, T., Gopnik, A., Schulz, L., & Danks, D. (2003). *Inferring hidden causes*. Presented at the annual meeting of the Cognitive Science Society, Boston.

Landauer, T. K. (1998). Learning and representing verbal meaning: The latent semantic analysis theory. *Current Directions in Psychological Science*, *7*, 161–164.

Landauer, T. K., Foltz, P., & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes*, *25*, 259–284.

Lockhart, K., Aw, J., & Essig, G. (2004, April 24). *Children's beliefs about the origins of positive and negative traits*. Poster presented at the biennial conference on Human Development, Washington, DC.

Lockhart, K. L., Chang, B., & Story, T. (2002). Young children's beliefs about the stability of traits: Protective optimism? *Child Development*, *73*, 1408–1430.

Lockhart, K. L., Nakashima, N., & Inagaki, K. (2004, August 8-13). *Japanese and American beliefs about the origins of positive and negative traits*. Paper presented at the International Congress of Psychology, Beijing, China.

Lockhart, K. L., Nakashima, N., Inagaki, K., & Keil, F. C. (2008). From ugly duckling to swan? Japanese and American beliefs about the stability and origins of traits. *Cognitive Development*, *23*, 155–179.

Lutz, D. J., & Keil, F. C. (2002). Early understanding of the division of cognitive labor. *Child Development*, *73*, 1073–1084.

Medin, D. L., & Waxman, S. R. (2007). Interpreting asymmetries of projection in children's inductive reasoning. In A. Feeney & E. Heit (Eds.), *Inductive reasoning* (pp. 55–80). New York: Cambridge University Press.

Miller, J. L., & Bartsch, K. (1997). Development of biological explanation: Are children vitalists? *Developmental Psychology*, *33*, 156–164.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, *92*, 289–316.

Nakashima, N., Lockhart, K., & Inagaki, K. (2003). *Japanese and American beliefs about the stability of traits*. Poster presented at the annual meeting of the Society for Research in Child Development, Tampa, FL.

Notaro, P. C., Gelman, S. A., & Zimmerman, M. A. (2001). Children's understanding of psychogenic bodily reactions. *Child Development*, *72*, 444–459.

Opfer, J. E., & Siegler, R. S. (2004). Revisiting preschoolers' living things concept: A microgenetic analysis of conceptual change in basic biology. *Cognitive Psychology*, *49*, 301–332.

Poulin-Dubois, D., & Shultz, T. R. (1990). Infants' concept of animacy. *Cognitive Development*, *11*, 19–36.

Rakison, D. H., & Poulin-Dubois, D. (2001). Developmental origin of the animate-inanimate distinction. *Psychological Bulletin*, *127*, 209–228.

Rochat, P., Morgan, R., & Carpenter, M. (1997). Young infants' sensitivity to movement information specifying social causality. *Cognitive Development*, *12*, 441–465.

Rosengren, K. S., Gelman, S. A., Kalish, C. W., & McCormick, M. (1991). As time goes by: Children's early understanding of growth in animals. *Child Development*, *62*, 1032–1320.

Ross, N., Medin, D. L., Coley, J. D., & Atran, S. (2003). Cultural and experiential differences in the development of folkbiological induction. *Cognitive Development*, *18*, 25–47.

Rozenblit, L. R., & Keil, F. C. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, *26*, 521–562.

Shulz, L. E., Bonawitz, E. B., & Griffiths, T. L. (2007). Can being scared cause tummy aches?: Naïve theories, ambiguous evidence, and preschooler's causal inferences. *Developmental Psychology*, *43*, 1124–1139.

Siegel, M. (1988). Children's knowledge of contagion and contamination as causes of illness. *Child Development*, *59*, 1353–1359.

Simons, D. J., & Keil, F. C. (1995). An abstract to concrete shift in the development of biological thought: The *insides* story. *Cognition*, *56*, 129–163.

Smith, M. C. (1978). Cognizing the behavior stream: The recognition of intentional action. *Child Development*, *49*, 736–743.

Solomon, G. E. A. (2002). Birth, kind and naïve biology. *Developmental Science*, *5*, 213–218.

Solomon, G. E. A., Johnson, S. C., Zaitchik, D., & Carey, S. (1996). Like father, like son: Young children's understanding of how and why offspring resemble their parents. *Child Development*, *67*, 151–171.

Sommerville, J. A. (2007). Detecting causal structure: The role of intervention in infants' understanding of psychological and physical causal relations. In A. Gopnik & L. Shulz (Eds.), *Causal learning: Psychology, and computation* (pp. 48–57). Oxford, UK: Oxford University Press.

Sommerville, J. A., & Woodward, A. L. (2005). Infants' sensitivity to the causal features of means-end support sequences in action and perception. *Infancy*, *8*, 119–145.

Spelke, E. S., Phillips, A., & Woodward, A. L. (1995). Infant's knowledge of object motion and human action. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 44–78). Oxford, UK: Clarendon.

Springer, K. (1992). Children's awareness of the biological implications of kinship. *Child Development*, *63*, 950–959.

Springer, K. (1996). Young children's understanding of the biological basis of parent–offspring relations. *Child Development*, *67*, 2841–2856.

Springer, K., & Ruckel, J. (1992). Early beliefs about the causes of illness: Evidence against immanent justice. *Cognitive Development*, *7*, 429–443.

Steyvers, M., Tenenbaum, J. B., Wagenmakers, E. J., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science*, *27*, 453–489.

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science*, *18*, 580–586.

Waxman, S. R., Medin, D. L., & Ross, N. (2007). Folkbiological reasoning from a cross-cultural developmental perspective: Early essentialist notions are shaped by cultural beliefs. *Developmental Psychology*, *43*, 294–308.

Wellman, H. M., & Gelman, S. A. (1992). Cognitive development: Foundational theories of core domains. *Annual Review of Psychology*, *43*, 337–375.

Wellman, H. M., & Gelman, S. A. (1998). Knowledge acquisition in foundational domains. In W. Damon (Ed.), *Handbook of child psychology* (Vol. 2, 5th ed., pp. 523–573). New York: Wiley.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford, UK: Oxford University Press.

Appendix A

Experiment 1: Human Biological and Psychological Processes

Biology

General health.

1. Why people's bones get stronger from drinking milk.*
2. Why people feel good when they exercise.
3. Why people should take vitamins.
4. How people get energy from food.*
5. How people build big muscles from protein.*
6. Why people breathe faster when exercising.
7. Why people need sleep to grow taller.
8. Why people need to drink lots of water.*
9. Why people get wrinkly skin when they're older.
10. Why people get tired if they do not eat enough.
11. Why people need to eat vegetables.
12. Why people are sometimes allergic to things.

Body breakdown/illness.

1. How people catch a cold.
2. Why people's skin bruises when they bump it.
3. Why people sometimes get tummy aches.
4. Why people get stuffy noses.
5. Why people's cuts scab over.*
6. Why people need more rest when they're sick.
7. Why people get cavities from eating too much candy.
8. How people's broken bones heal.*
9. Why people get better when they take medicine.*
10. Why people itch from chicken pox.
11. Why people get blisters if their shoes don't fit.
12. How people get sick from germs.

Psychology

Cognitive.

1. Why people think clearer when it is quiet.*
2. Why people find it hard to do math in their head.
3. Why people use their fingers to count to ten.
4. Why people sometimes think out loud.*
5. How people remember phone numbers.*
6. Why people sometimes remember things wrong.
7. Why people find puzzles hard to do.
8. Why people sometimes forget things.
9. How people solve riddles.*
10. Why people learn the alphabet easier if they sing it many times.
11. How people imagine things in their mind.
12. Why people learn languages easier when they're younger.

Personality/motivational.

1. Why people feel good when they do well in school.*
2. Why people act shy.
3. Why people like to learn new things.
4. Why people sometimes talk a lot.*
5. Why people like to be alone sometimes.
6. Why people are bullies.
7. Why people are messy.*
8. Why people work harder if they are rewarded.
9. Why people like to do things they're good at.
10. Why people are bossy.*
11. Why people get mad when they don't get what they want.
12. Why people like different things from one another.

Note. Asterisks indicate the items used in the questions. All other items were answer options.

Appendix B

Experiment 2: Animal Biological and Psychological Processes

Biology

Anatomy.

1. Why cats have blood in their bodies.
2. Why cats grow sharp claws on their paws.
3. Why horses' stomachs are so big.*
4. Why dogs get tummy aches.*
5. How birds have beaks and not teeth.
6. How birds grow feathers on their wings.
7. Why fish have scales all over their body.*
8. Why elephants' skin looks so wrinkly.
9. Why elephants grow long trunks.
10. Why camels have long eyelashes.
11. Why camels' humps are full of fat.
12. Why monkeys have long tails.*

Physiology

1. Why cats need to eat meat to stay healthy.*
2. Why horses sweat when it's hot outside.
3. Why horses need more sleep when they're sick.
4. How dogs shed when it's hot outside.
5. How dogs breathe faster when exercising.
6. How birds heal when they get hurt.*
7. How fish breathe underwater.
8. How fish first grow in eggs.
9. Why elephants have to drink a lot of water.*
10. Why camels do not need to eat food very often.*
11. How monkeys need sleep to grow bigger.
12. How monkeys need milk when they're young.

Psychology

Cognitive.

1. How cats remember their way home.*
2. How horses remember people who ride them.*
3. How dogs learn tricks like rolling over or lifting a paw.
4. How dogs know who their owner is.
5. How birds learn different songs to sing.
6. How birds know where to fly in the winter.
7. How fish decide what other fish to swim with.
8. How fish know where the surface of the water is.
9. Why elephants have very good memories.*
10. How camels learn to carry people on their backs.*
11. How monkeys learn to use tools.
12. How monkeys learn to count things.

Emotional.

1. Why cats have fun playing with yarn.
2. Why cats like to be alone a lot of the time.
3. Why horses get scared when they hear a loud noise.
4. Why horses have a lot of fun rolling around on the ground.
5. Why dogs really enjoy being petted.*
6. Why birds feel scared during thunderstorms.*
7. Why fish feel scared when there's lots of splashing.*
8. Why elephants fight when they're angry.
9. Why elephants like to be near their mothers.
10. Why camels like traveling with other camels.
11. Why camels get mad easily.
12. Why monkeys are curious and like to explore.*

Note. Asterisks indicate the items used in the questions. All other items were answer options.

Child Development, January/February 2010, Volume 81, Number 1, Pages 410–426

Early Educational Intervention, Early Cumulative Risk, and the Early Home Environment as Predictors of Young Adult Outcomes Within a High-Risk Sample

Elizabeth P. Pungello, Kirsten Kainz,
Margaret Burchinal, Barbara H. Wasik,
and Joseph J. Sparling
University of North Carolina at Chapel Hill

Craig T. Ramey
Georgetown University

Frances A. Campbell
University of North Carolina at Chapel Hill

The extent to which early educational intervention, early cumulative risk, and the early home environment were associated with young adult outcomes was investigated in a sample of 139 young adults (age 21) from high-risk families enrolled in randomized trials of early intervention. Positive effects of treatment were found for education attainment, attending college, and skilled employment; negative effects of risk were found for education attainment, graduating high school, being employed, and avoiding teen parenthood. The home mediated the effects of risk for graduating high school, but not being employed for teen parenthood. Evidence for moderated mediation was found for educational attainment; the home mediated the association between risk and educational attainment for the control group, but not the treated group.

Numerous studies have documented the negative impact of multiple social risk factors on children's development from early childhood through adolescence (e.g., Sameroff, Seifer, Barocas, Zax, & Greenspan, 1987; Trentacosta et al., 2008). High-quality early educational programs are widely viewed by researchers, parents, and policy makers as a means of enhancing cognitive and social skills for young children exposed to such risk factors (Heckman, 2006). Enhancing early development is expected to lead to more positive educational, occupational, and social outcomes in adulthood. Findings from the Abecedarian Project, a study of intensive early educational intervention delivered to high-risk children in a child-care setting, support this expectation by demonstrating that individuals randomly assigned to early educational treatment, when compared with those assigned to the control group, maintained cognitive gains and showed educational and occupational benefits into young adulthood (Campbell, Pungello, Miller-Johnson, Burchinal, & Ramey, 2001; Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002). Long-term educational and occupational benefits were replicated in a subsequent randomized study of early education, the Carolina Approach to Responsive Education (CARE; Campbell et al., 2008). The present study adds to the literature concerning the long-term effects of educational intervention and early risk by examining: (a) the extent to which exposure to multiple social risk factors in early childhood predicts outcomes in young adulthood over and above the effects of early educational intervention within a high-risk sample, and whether early risk and early intervention interact to influence adult outcomes, and (b) whether such distal risk factors are

---

This research was supported in part by grants from the National Institutes of Health (R01 HD040817), the Maternal and Child Health Bureau (6 R40 MC 00254, 6 R40 00067), the Office of Educational Research and Improvement (R306F960201), and the David and Lucille Packard Foundation. The authors also wish to thank Katherine Polk, the study's project coordinator, and Carrie Bynum, the study's family coordinator, for their tireless efforts, without which this work could not have been completed. Our extreme gratitude also goes out to the families who have participated in this work over these many years.

Correspondence concerning this article should be addressed to Elizabeth P. Pungello, Frank Porter Graham Child Development Institute, Campus Box 3270, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-8180. Electronic mail may be sent to liz_pungello@unc.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8101-0027

associated with adult outcomes through proximal processes associated with the quality of the early home environment, and whether early intervention moderates the effects of the proximal processes in the early home on young adult outcomes.

Early Educational Intervention and Early Cumulative Risk

Head Start, public prekindergarten, and, to a lesser extent, subsidized child-care programs have been funded by local, state, and federal governments in an effort to enhance the early development of children raised in poverty to help them overcome their increased risk of academic failure, unemployment, teenage parenthood, and criminal behavior as young adults. These efforts have been undergirded by results from randomized experiments demonstrating enhanced outcomes associated with high-quality preschools or educational intervention in child-care settings targeting poor children (Berrueta-Clement, Schweinhart, Barnett, Epstein, & Weikart, 1984; Campbell et al., 2001, 2008; Hubbs-Tait et al., 2002) and observational studies suggesting that quality child care enhances children's development (see Vandell, 2004, for a review). Both the Abecedarian and CARE studies found positive effects for young adult education and occupational outcomes (Campbell et al., 2002; Campbell et al., 2008). More specifically, compared to controls, those who received the early educational intervention attained more years of education, were more likely to attend a 4-year college or university, and were more likely to have obtained skilled employment. However, no significant differences were found in either sample for high school graduation or employment rates. In addition, those treated in the Abecedarian Project were less likely to be teen parents compared to controls, although this was not replicated in the CARE sample. Neither the Abecedarian nor CARE projects found reductions in criminal behavior, but such effects were found for the Perry Preschool Study (Schweinhart, Barnes, & Weikart, 1993).

The multiple-risk model proposed by Rutter (1979) and Garmezy, Masten, and Tellegen (1984) posits that developmental outcomes are a function of individual responses to risk factors. This model focuses on pathways to competence in the context of adversity (Masten et al., 1999) and emphasizes identifying "protective factors" that weaken the link between adversity and child outcomes and promote "successful adaptation despite challenging or threatening circumstances" (Masten, Best, & Garmezy, 1990, p. 426). The multiple-risk approach focuses on risk composites that describe the extent of exposure to various factors based on the recognition that distal indices such as poverty, single parenthood, large households, low parental education, unemployment, low-income communities and poor-quality schools, and more proximal measures, such as maternal depression and lack of social support, tend to cluster in the same individual (Masten et al., 1995). Accounting for these correlated constraints through multiple or cumulative risk indices may provide better theoretical and empirical models of how exposure to negative factors impacts children's development than does examining any single individual risk factor or examining them in an additive manner (Sameroff & MacKenzie, 2003).

Several studies have demonstrated that high scores on multiple risk indices are negatively related to cognitive, language, and socioemotional outcomes in early childhood, middle childhood, and adolescence (Brody & Flor, 1998; Brody, Kim, & Murry, 2003; Burchinal, Roberts, Hooper, & Zeisel, 2000; Burchinal, Roberts, Zeisel, Hennon, & Hooper, 2006; Evans, 2003; Forehand, Biggar, & Kochick, 1998; Gerard & Buehler, 2004; Gutman, Sameroff, & Eccles, 2002; Gutman, Sameroff, & Cole, 2003; Hooper, Burchinal, Roberts, Zeisel, & Neebe, 1998; Jones, Forehand, Brody, & Armistead, 2002; Krishnakumar & Black, 2002; Liaw & Brooks-Gunn, 1994; Linver, Brooks-Gunn, & Kohen, 2002; Luster & McAdoo, 1994; Prelow & Loukas, 2003; Pungello, Kupersmidt, Burchinal, & Patterson, 1996; Sameroff, Seifer, Baldwin, & Baldwin, 1993; Sameroff et al., 1987; Trentacosta et al., 2008). Less is known, however, about the very long term effects of early exposure to multiple risk factors. Young adulthood, spanning ages 18–25, is a time termed "emerging adulthood" by Arnett (2000), a life stage encompassing the transition between separation from the family of origin and becoming self-supporting. Previous research has linked individual early risk factors to young adult education outcomes, employment, teen parenthood, and criminal behavior (Aquilino, 1996; Ensiminger & Slusarcick, 1992; Gest, Mahoney, & Cairns, 1999; Haurin, 1992; Jaffee, 2002; Jimerson, Egeland, Sroufe, & Carlson, 2000; Xie, Cairns, & Cairns, 2001; Zill, Morrison, & Coiro, 1993), but to date, less research has examined whether exposure to multiple social risk factors during early childhood predicts adult outcomes. An exception is a study of over 30,000 individuals in Britain in which Schoon et al. (2002) found increased socioeconomic risk to be associated with lower academic attainment through adolescence and lower social class attainment in adulthood.

Studies examining the effects of cumulative risk often include income and ethnicity among the multiple risk indices, yet investigating how cumulative risk affects outcomes within samples limited to such individuals is needed given their increased likelihood of experiencing multiple stressors. Children in poverty are more likely to experience both social risk (e.g., family disruption) and environmental risk (e.g., pollution; see review by Evans, 2004). African American children often experience racism in addition (Luster & McAdoo, 1994). Thus, ''sociocultural forces such as poverty and racism tend to allocate risk disproportionately… to subsets of the population such as poor and ethnic minorities'' (Evans, 2003, p. 924). Using heterogeneous samples, researchers have investigated the effects of additional risk over and above those associated with income and ethnicity, finding somewhat inconsistent results depending on the additional risk included (e.g., stressful life events, Pungello et al., 1996; low birth weight, Liaw & Brooks-Gunn, 1994). Others have examined the effects of additional risk within high-risk samples (Burchinal, Roberts, et al., 2000; Hooper et al., 1998; Shaw, Winslow, Owens, & Hood, 1998; Burchinal et al., 2006; Krishnakumar & Black, 2002; Brody et al., 2003; Gutman et al., 2002; Prelow & Loukas, 2003), finding negative associations between such risk and children's outcomes. However, although some longitudinal work has been carried out (Brody et al., 2003; Burchinal, Roberts, et al., 2000; Burchinal et al., 2006), most of these studies have investigated short-term links between risk and outcomes.

In addition to investigating the main effects of early education in a child-care setting and early cumulative risk, researchers have tested for interactions between early care experiences and early risk. Both compensatory effects (i.e., children with higher risk benefited more than children with lower risk) and leveraging effects (i.e., greater effects found for children with less risk than for those with more risk) have been found. Hubbs-Tait et al. (2002) found children with high family risk benefited more by Head Start attendance than those at low risk when predicting language scores. In contrast, Liaw and Brooks-Gunn (1994) examining the effects of the Infant Health and Development Program found that for children raised in poverty, the intervention had a greater effect for children with less risk than for children who experienced more risk.

The first goal of this study was to further the understanding of the effects of early intervention and early cumulative risk by: (a) investigating the associations between exposure to multiple social risk factors in early childhood and young adult outcomes within a sample drawn entirely from low-income families, almost all of whom were African American, over and above the effects of early educational intervention, and (b) examining whether early intervention and early risk interact to influence outcomes in young adulthood.

Moderated Mediation: The Early Home Environment

A number of risk indices have been constructed to examine how exposure to risk negatively affects developmental outcomes. Although both distal (e.g., poverty, unsafe neighborhoods, maternal education) and proximal (e.g., quality of the home environment) factors have often been included in risk indices (Barocas et al., 1991; Hooper et al., 1998; Luster & McAdoo, 1994; Prelow & Loukas, 2003; Sameroff et al., 1993), more recent studies have focused on how exposure to distal risk factors affects development via proximal processes (e.g., Brody et al., 2003; Trentacosta et al., 2008). Based upon ecological theory (Bronfenbrenner & Morris, 1998), researchers have hypothesized that the stress and lack of opportunities associated with poverty, low education, and large households may significantly diminish the family's psychological strengths, resulting in less responsive care by parents (e.g., Conger & Elder, 1994). Thus, not only do risk factors tend to cluster together, but the co-occurrence of multiple risks may overwhelm the family's ability to cope and provide positive parenting to the child (Sameroff et al., 1987).

A number of explanatory models have been tested in previous work, including additive (testing for independent effects of each variable), cumulative (testing the effect of a total risk score), interactive (testing if proximal variables moderate the effects of distal variables), and mediational (testing if the effects of distal variables are mediated through their effects on proximal variables), with the mediation model receiving the greatest support (e.g., Jones et al., 2002; Krishnakumar & Black, 2002). For example, proximal factors such as the quality of the home environment mediated distal risk factors in predicting cognitive outcomes prior to the age of 6 years (Barocas et al., 1991; Krishnakumar & Black, 2002), externalizing and internalizing problems among high-risk children in early childhood (Trentacosta et al., 2008), academic and social outcomes among

African American children in Grades 1–3 (Burchinal et al., 2006), and social outcomes for African American children 7–15 years of age (Brody et al., 2003; Jones et al., 2002). Thus, the weight of evidence from these and other prior studies supports the hypothesis that distal risk factors, such as poverty, maternal education, and family structure, predict child outcomes through effects on proximal processes (Guo & Harris, 2000; Guo & VanWey, 1999; Luster & McAdoo, 1994; National Institute of Child Health and Human Development Early Child Care Research Network, 2005). However, less is known about the duration of these effects, that is, the extent to which risk early in the life span may influence the quality of the home environment in early childhood, which may in turn affect outcomes across many years, even into young adulthood.

In addition, researchers have examined whether early care experiences may moderate the effects of the early home environment, finding both compensatory (Bradley, Corwyn, Burchinal, McAdoo, & Garcia-Coll, 2001) and leveraging (Bryant, Burchinal, Lau, & Sparling, 1994) effects. This is similar to findings from the studies examining the interaction between early care experiences and early cumulative risk described earlier.

The second overall goal of this study was to further understanding of the role of the early home environment by: (a) examining whether the early home environment mediates the effects of early cumulative risk on young adult outcomes and (b) investigating whether early educational intervention in a full-time child-care setting moderates the effects of the early home environment on young adult outcomes.

Present Study

This study used data from the Abecedarian Project (Ramey & Campbell, 1984) and CARE (Wasik, Ramey, Bryant, & Sparling, 1990) to examine the very long term effects of early educational intervention in a child-care setting, early multiple risk exposure, and the early home environment. These consecutive longitudinal programs have followed participating individuals from infancy through young adulthood. Early risk factors and the home environment were assessed during program implementation from infancy through age 5 (kindergarten entry), and the data on early circumstances were entered into predictive models to explain educational outcomes, employment outcomes, teen parenthood, and criminal behavior when these persons were 21–25 years of age.

Our hypotheses were as follows. Concerning the first overall goal of the study, prior analyses with these samples have demonstrated that those who received the early educational intervention showed benefits in terms of education and vocational outcomes (Campbell et al., 2002; Campbell et al., 2008). We hypothesized that within this high-risk sample, over and above any found effects of treatment, early cumulative risk would be negatively associated with young adult educational outcomes (education attainment in general, graduating high school specifically, and attending college specifically), employment outcomes (being employed, obtaining skilled employment), avoidance of teen parenthood, and avoidance of criminal behavior (i.e., individuals exposed to more risk would be more likely to be convicted of a misdemeanor, more likely to be convicted of a felony, and more likely to use illegal drugs). We also hypothesized that early intervention would moderate the effects of risk such that the effects of increased risk would be weaker for those who received the intervention than for those who did not. Concerning the second goal of the study, we hypothesized that the early home environment would mediate any found effects for early risk and that early educational intervention would moderate the effects of the early home environment such that the effects of a poor-quality home environment would be weaker for those who received treatment compared to those who did not.

Method

Participants

Participants were 139 young adults enrolled as infants in one of the two consecutive trials of early educational intervention. The Abecedarian study enrolled four cohorts of infants born between 1972 and 1977, and CARE added two more born between 1978 and 1980. To determine eligibility for both samples, families were screened with the same High Risk Index (Ramey & Smith, 1977) composed of sociodemographic factors associated with delays in cognitive development and educational failure. Among the factors were the educational levels of parents, family income, family structure, evidence of cognitive delays or academic failure in other family members, and the use of welfare funds to meet basic needs. Scores on the index were weighted and summed to measure the degree of risk for each family; to qualify, a score of 11 or higher was required.

The original Abecedarian sample consisted of 111 infants; CARE consisted of 66. Ninety-four percent of the original high-risk families in the combined sample identified themselves as African American (the remaining 6% were Euro-American), and 53% of the infants were male. In the Abecedarian study, infants were randomly assigned to two groups: full-time educational treatment in a child-care setting or control. In the CARE study, infants were randomly assigned to three groups: full-time educational treatment in a child-care setting plus family education, family education only, or control. Table 1 summarizes for both samples the numbers of male and female study participants for whom data were collected in young adulthood. Attrition across the two studies was only 6%. At the time of the young adult follow-up, the individuals ranged in age from 20 to 25 years. For the Abecedarian study, the mean age at data collection was 21.2 years (*SD* = 0.60 years); for CARE, the mean age was 22.5 years (*SD* = 0.71 years).

Given that one of the main interests in this study was to learn the degree to which the early childhood educational treatment in a child-care setting influenced later outcomes in light of early risk and the quality of the early home environment, the present analyses were confined to individuals who received early educational treatment within the child-care setting (whether or not they also received family education) and those randomly assigned to the control groups. Thus, the treated group consisted of the Abecedarian treated group (*n* = 53), and the CARE treated group (*n* = 14). The control group consisted of the Abecedarian control group (*n* = 51), and the CARE control group (*n* = 21). The CARE group that received only the family education (*n* = 25 families) was not included in these analyses (see Wasik et al., 1990, for comparisons of early childhood outcomes for three CARE groups). Thus, the total sample size for the present analyses was 139 (73 males and 66 females; 67 who received early educational treatment and 72 controls). Analyses of young adult outcomes for the Abecedarian and CARE projects have been based on an "intent-to-treat" model in which all individuals were classified according to their random group assignment in infancy. For specific outcomes, the numbers vary slightly depending upon the few young adults who chose not to disclose certain information.

Table 1
Number of Abecedarian and CARE Study Participants Interviewed as Young Adults

	Abecedarian		CARE		
Group	Males	Females	Males	Females	Total
Early educational treatment	28	25	9	5	67
Control	23	28	13	8	72
Family education only			15	10	25
Total	51	53	37	23	164

Note. The child-care treatment group in the Carolina Approach to Responsive Education (CARE) study received family education as well.

Although the Abecedarian and CARE studies comprised separate samples, several factors provide confidence that they can be combined into one sample. First, the children in both the Abecedarian and CARE samples were recruited in the same manner from the same communities, and the center-based treatment in both samples was essentially the same (the early educational intervention staff was relatively stable, they were supervised by the same investigators, and the site and the curriculum were the same). Further, earlier analyses examining the effects of treatment on cognitive outcomes through age 8 found treatment and not sample to be substantively related to outcomes (with IQ differences between the samples being found at only 2 of the 10 assessment points) and no Treatment × Sample interaction (Burchinal, Campbell, Bryant, Wasik, & Ramey, 1997). Prior analyses for age 21 outcomes compared 12 basic background factors (maternal age, education, marital status, teen parent, ethnicity, etc.) across the two study samples and found only one significant difference: Mothers of infants enrolled in CARE had more years of education at the time of the target child's birth (Campbell et al., 2008). These prior young adult analyses also compared the effects of treatment in the samples and found no evidence for a larger treatment effect in the CARE sample (i.e., for individuals who received both the child-care treatment and the family education component) than the Abecedarian sample (i.e., those who received only the child-care treatment) for long-term educational outcomes (Campbell et al., 2008). Finally, in analyses for the current study (described in the following), logistic regression on each of the individual risk factors that comprised the cumulative risk index employed in this study and a general linear model for the risk total score found no evidence that individual risk items or the total scores differed across samples.

Procedures

Treatment and control groups. The Abecedarian and CARE treatment involved early educational intervention beginning in infancy and lasting until the child started kindergarten. Age at entry ranged from 3 to 22 weeks, with a mean age of 9 weeks ($SD = 4.6$ weeks). The intervention was delivered in a full-time child-care facility housed in a university-based research center. Descriptions of the educational program and early childhood outcomes are provided in a number of earlier publications (i.e., McGinness & Ramey, 1981; Sparling, 1989; Sparling & Lewis, 1978, 1984; Ramey & Campbell, 1984; Wasik et al., 1990).

Longitudinal data collection. From infancy to age 5 years (i.e., early childhood), demographic circumstances for each child's family were repeatedly assessed. Data were collected at the point of study admission and at least annually thereafter to track changes in the child's family composition, parental education levels, and stability of living circumstances. Annual home visits were also made during which the educational atmosphere of the child's home environment was assessed. The focus of the present work is on how events in the first 5 years of life were related to long-term outcomes as represented in accomplishments assessed in the young adulthood follow-up.

Young adult follow-up. Young adults were contacted by letter and invited to enroll in this phase of the study. Assessors were advanced graduate students in clinical or school psychology. All were unaware of the participants' early treatment histories. Two of the assessors were African American and one was European American for the Abecedarian follow-up; the three assessors were European American for the CARE follow-up.

Measures

Early risk. The following variables were selected from data collected between infancy and 54 months to create the present cumulative risk index: teen mother when the participant was born (mother younger than 18), mother's educational level at birth less than high school graduate, parents not married at some point during this time period, participant did not live with the mother at some point during this period, large family size (participant had two or more siblings by age 54 months), and high mobility (the family made three or more moves during this time period). These factors were selected based upon prior studies examining the effects of early risk factors (Aquilino, 1996; Burchinal, Roberts, et al., 2000; Chase-Lansdale, Gordon, Brooks-Gunn, & Klebanov, 1997; Deater-Decker, Dodge, Bates, & Pettit, 1998; Dubow & Luster, 1990; Ensiminger & Slusarcick, 1992; Furstenberg, Brooks-Gunn, & Chase-Lansdale, 1989; Gutman et al., 2002; Hooper et al., 1998; Luster & McAdoo, 1994; McLanahan, 1997; Murry, Bynum, Brody, Willert, & Stephens, 2001; Prelow & Loukas, 2003; Sameroff et al., 1993; Sugawara, 1991; Williams, Anderson, McGee, & Silva, 1990; Zill et al., 1993). The cumulative risk score was derived by summing up 1 point for each of the factors that pertained to the family during the child's first 5 years. Thus, early risk scores could range from 0 to 6.

The early home environment. Abecedarian and CARE families were visited at home when study children were 6, 18, 30, 42, and 54 months of age. Based on observations and questions during the visits, the age-appropriate version of the Home Observation for Measurement of the Environment (HOME; Bradley & Caldwell, 1979; Caldwell & Bradley, 1984) was scored. Versions of the HOME suitable for infants/toddlers and preschool children aged 3–6 covered factors such as the affective quality of the parent (mother)–child interaction, the toys and educational materials provided, the parent's support for the child's learning, the stability of the family's routines, and the variety and breadth of stimulation made available to the child. Caldwell and Bradley (1984) reported internal consistency reliability of $r = .89$ for the infant/toddler version of the HOME, and $r = .93$ for the preschool version (Bradley, Caldwell, Rock, Hamrick, & Harris, 1988). Bradley (1992) found comparable factor structures and predictive validity for European American and African American samples. In the present sample, alpha levels for the HOME ranged from .75 at 6 months to .89 at 54 months, with an overall average alpha of .82.

The infant/toddler and preschool versions of the HOME differ in the number of items but are highly correlated ($.43 < r < .72$). Accordingly, an across-time composite HOME score was computed as the mean of the percent of items passed at 6, 18, 30, 42, and 54 months. This composite HOME score ranged from 0.40 to 0.94 ($M = 0.69$ and $SD = 0.11$). Thus, on average, 69% of the items were passed over time across the families in the sample.

Young adult outcomes. Outcomes for the analyses presented here include educational attainment, employment, teen parenthood, and criminal behavior. These data were collected through means of a semistructured interview covering these and other basic demographic factors. The use of illegal drugs was measured by self-scored answers to questions contained in the Behavioral Risk Factor Surveillance System (Kolbe, 1990).

Educational attainment was operationalized in three ways. First, a continuous measure of education was created based on the number of years associated with the final degree obtained. The score was the highest grade completed if the participant did not graduate from high school or obtain a general equivalency diploma (GED), the score was 12 if the participant had graduated from high school or obtained a GED, 14 if the participant had completed some college or obtained an associate's degree, and 16 if the participant had obtained a bachelor's degree. To address policy questions, two categorical variables were also created: high school graduate (*yes–no*) and ever attended a 4-year college or university (*yes–no*).

Two binary employment outcomes were examined. First, whether or not the participant was employed in any capacity at the time of the young adult interview was analyzed. Second, positions were coded according to the Hollingshead Index of Social Class (Hollingshead, n.d.) and whether or not the participant was employed in a skilled labor occupation or higher (defined as Hollingshead rating of 4 or higher) was coded.

Teen parenthood was defined as having a first child before age 18. Criminal behavior was self-reported: whether or not the participant had been convicted of a misdemeanor, convicted of a felony, or reported any illegal drug use.

Analysis Strategy

Two sets of analyses were conducted to address the two goals of the study. First, the simultaneous effects of treatment and risk were investigated using general linear modeling (in the models predicting to the continuous variable for education attainment) and logistic regression techniques (in the models predicting to the dichotomous outcomes: high school graduate, ever attended college, currently employed, currently employed in a skilled job, teen parenthood, ever convicted of a misdemeanor, ever convicted of a felony, and use of illegal drugs) in SAS version 9.1 (Statistical Analysis Software, Cary, NC). For each outcome, two predictive models were run. The first tested for the main effects of early intervention and early risk; the second added an intervention by risk interaction term to the model.

In the second set of analyses, young adult developmental status was conceptualized within a moderated mediation framework, hypothesizing that the early home environment mediated the effect of early risk on young adult outcomes and that early educational intervention status moderated the relation between the home environment and young adult outcomes (see Figure 1). Thus, for each outcome where significant effects of early risk above and beyond treatment were found in the first set of analyses, the moderated mediation hypothesis was investigated using Mplus version 5 (Muthén & Muthén, 1998–2007). Mplus provides several advantages for this analysis in that it: (a) allows for the proper specification and analysis of binary and continuous outcomes, (b) estimates bootstrapped standard errors and confidence intervals for each model coefficient, and (c) addresses missing data through a full information maximum likelihood (FIML) technique. Following recommendations by Preacher, Rucker, and Hayes (2007), the moderated mediation path model included the early risk index, early home environment, early childhood educational treatment status, and a Home Environment × Treatment interaction term as predictors of young adult outcomes (see Figure 2). Additionally, the home environment was regressed on the child risk index. Each predictor was mean-centered for analysis and for the creation of the interaction term. The analysis plan focused on assessing the significance of simultaneous mediation and moderation within the path models. Significant mediation was determined by a nonzero product term of the a_1 and b_1 paths, controlling for all other variables in the model; this product term is referred to as the indirect effect or mediated effect (MacKinnon, 2008). The term indirect effect is used from this point forward. If the path between the interaction term and outcome was significant for a specific young adult outcome, post hoc evaluation of simple indirect effects within treatment levels was conducted (Tein, Sandler, MacKinnon, & Wolchik, 2004).

Results

Descriptive Findings

The number of early risk factors experienced by study participants ranged from 0 to 4 ($M = 2.31$, $SD = 1.10$). The mean cumulative risk scores were

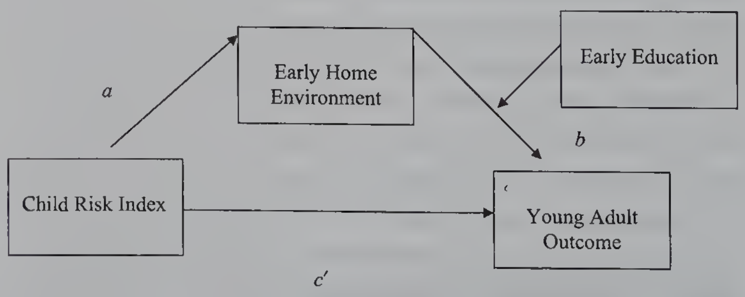


Figure 1. Conceptualization of moderated mediation.

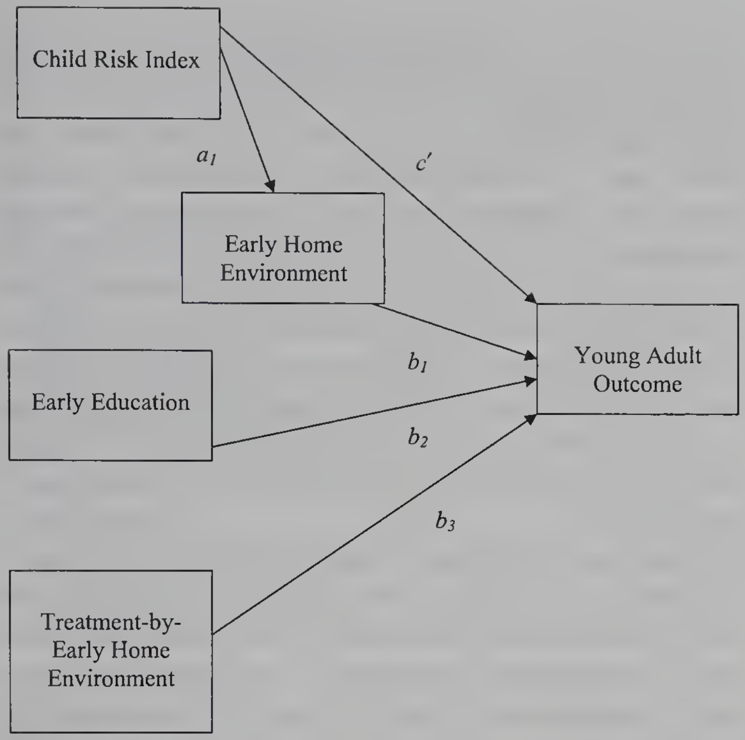


Figure 2. Moderated mediation path model specification.

2.32 ($SD = 1.06$) for the treated group and 2.38 ($SD = 1.18$) for the control group. The mean early home environment score for the treated group was 0.68 ($SD = 0.091$) and for the control group was 0.66 ($SD = 0.10$). There were no significant differences between the treated and control groups for any risk item, the total risk score, or the home environment score.

Table 2 provides the percentages of the treatment and control groups who experienced each risk factor and each of the binary young adult outcomes. Logistic regressions for each of the binary outcomes with only treatment as the predictor found significant differences between the treated and control groups for having attended college, $B = 1.31$ (0.43), $p < .01$, and skilled employment, $B = 0.98$ (0.37), $p < .01$. For the continuous educational outcome, the mean for the total sample was 12.31 ($SD = 1.64$), with $M = 12.77$ ($SD = 1.45$) for the treated group and $M = 11.88$ ($SD = 1.69$) for the control group; a general linear model with treatment alone as a predictor found a significant difference between the treated and control groups for this outcome, $B = 0.90$ (0.27), $p < .01$. The correlation between educational attainment and cumulative risk was $r = -.20$ ($p < .05$), and the correlation between educational level and the early home environment was $r = .36$ ($p < .0001$).

Research Goal 1: Prediction Models

Table 3 presents raw regression coefficients and standard errors for the linear education attainment outcome and odds ratios (ORs) and confidence intervals (CIs) for dichotomous outcomes. The findings indicate that when the effects of treatment and

Table 2

Descriptive Statistics for Individual Risk Factors and Binary Outcomes by Treatment Group

	Group					
	Treated		Control		Total	
	n	% of group	*n*	% of group	*N*	% of sample
Risk factors						
Teen mother	15	22.7	21	29.2	36	26.1
Mother less than high school graduate	37	56.1	46	63.9	83	60.1
Mother not always married	59	89.4	61	84.7	120	87.0
Mother out of home	10	15.5	8	11.1	18	13.0
Large family size	13	19.7	17	23.6	30	21.7
Frequent moves	20	30.3	18	25.0	38	27.5
Outcomes						
High school graduate	50	74.6	48	72.2	98	71.0
Attended college	25	37.9	10	14.8	35	25.6
Employed as young adult	45	67.2	40	56.3	85	61.6
Skilled employment	30	46.2	17	24.3	47	34.8
Teen parent	8	11.9	15	20.8	23	16.6
Misdemeanor	8	12.3	14	19.4	22	16.1
Felony	6	9.2	8	11.1	14	10.2
Used illegal drugs	44	67.7	50	69.4	94	68.6

risk on young adult outcomes were simultaneously estimated, positive treatment effects were found for educational attainment, the likelihood of attending college, and the likelihood of having obtained skilled employment. Risk was negatively associated with educational attainment, the likelihood of high school graduation, the likelihood of being employed in any capacity, and the likelihood of avoiding teen parenthood. Effect sizes indicate that, controlling for risk, treated participants obtained 0.87 more years of education, were 3.82 times more likely to have attended some college, and were 2.69 times more likely to have obtained skilled employment compared to control participants. Controlling for treatment, each additional risk factor was associated with 0.28 fewer years of education, being 1.45 times less likely to graduate from high school, 1.61 times less likely to be employed, and 1.78 times more likely to be a teen parent (for ORs less than one reported in the table, the inverse of the reported value is used to describe the odds). Neither treatment nor risk appeared to have a significant effect on self-reported involvement in crime.

In the second set of models that tested for a treatment by risk interaction in addition to simultaneously estimating the main effects of treatment and risk, no evidence was found to suggest treatment moderated the associations between risk and any of the young adult outcomes.

Table 3

Young Adult Outcomes as a Function of Early Educational Intervention and Early Cumulative Risk

	Model 1			
	Treat		Risk	
	B	*SE*	*B*	*SE*
Education attainment	0.87**	0.27	−0.28*	0.12
	OR	CI	OR	CI
High school graduate	1.30	0.61, 2.77	0.69*	0.49, 0.97
Employed	1.53	0.75, 3.13	0.62**	0.44, 0.87
Teen parent	0.53	0.20, 1.38	1.78*	1.13, 2.80
Any college	3.82**	1.66, 8.80	0.91	0.63, 1.31
Skilled employment	2.69**	1.29, 5.60	0.88	0.63, 1.22
Felony	0.86	0.28, 2.65	1.60	0.94, 2.75
Misdemeanor	0.60	0.23, 1.56	1.51	0.98, 2.33
Illegal drugs	0.90	0.44, 1.85	0.95	0.69, 1.31

Note. Values for treatment and risk were centered for analysis.
$^{*}p < .05$. $^{**}p < .01$.

Research Goal 2: Moderated Mediation

The moderated mediation hypotheses was then tested for each outcome where a main effect for risk was found in the prediction models described earlier (i.e., educational attainment, graduated high school, being employed, and teen parenthood). Table 4 (for the educational outcomes) and Table 5 (for employment) provide the path coefficients, bias-corrected standard errors, and 95% CIs for the model path. In the final row of each table are the coefficient for the indirect effect, along with its bias-corrected standard error and CI.

Education attainment. Table 4 shows that early educational treatment, the home environment, and the Treatment × Home interaction significantly predicted education attainment. Additionally, there was a significant indirect effect of risk on educational attainment as mediated by the home environment. This indirect effect is conditional because the Treatment × Home interaction term was significant. Consequently, we followed up this finding with post hoc analyses of simple indirect effects by estimating the mediation model separately for the control and treated samples. These post hoc tests of simple indirect effects indicated that the home environment mediated the relation between risk and education attainment for participants in the control group (effect = −0.36, *SE* = 0.13, CI = −76, −0.11) but not for the treated group (effect = −0.04, *SE* = 0.05, CI = −21, 0.04). That is, for the control group, increases in education attainment were associated with increases in HOME scores ($B = 0.83$, $SE = 0.23$, $p < .000$), but not with child risk levels ($B = 0.01$, $SE = 0.14$, $p = .96$). For the treated group, education attainment was significantly related neither to HOME scores ($B = 0.19$, $SE = 0.19$, $p = .31$) nor to child risk levels ($B = -0.14$, $SE = 0.17$, $p = .41$).

Figure 3 depicts the Treatment × Home interaction effect on education attainment. For children with lower quality home environment scores (1 *SD* below the mean), those in the treatment group attained on average 2 more years of education than did children in the control group ($p < .001$). Treatment was not associated with a significant increase in education attainment for children with average-quality home environments ($p = .25$) or higher quality home environments (1 *SD* above the mean; $p = .24$).

High school graduation. As seen in Table 4, only the home environment accounted for high school graduation in this sample. The home environment, in turn, was predicted by children's risk index.

Table 4
Moderated Mediation: Educational Outcomes

	Educational attainment			High school graduation		
	Coefficient	*SE*	95% CI	Coefficient	*SE*	95% CI
Effect on outcome						
Education treatment	.68**	0.24	0.22, 1.16	.06	0.26	−0.43, 0.62
Risk	−.06	0.11	−0.27, 0.17	−.09	0.12	−0.32, 0.13
Home environment	.51**	0.15	0.22, 0.79	.30*	0.12	0.05, 0.51
Treatment × Home	−.57*	0.27	−1.02, −0.02	−.41	0.26	−0.95, 0.05
Effect on home environment						
Risk	−.33***	0.07	−0.48, −0.22	−.33***	0.06	−0.49, −0.23
Indirect effect						
Indirect effect	−.17**	0.06	−0.31, −0.06	−.10*	0.05	−0.21, −0.02

**p* < .05. ***p* < .01. ****p* < .001.

Thus, the effect of risk on high school graduation was significantly mediated by the home environment. However, there was no evidence of moderated mediation.

Employment. As seen in Table 5, as children's risk levels increased, their likelihood of being employed at the time of the young adult interview decreased, but no evidence that this risk was mediated through its effects on the early home environment was found. Conversely, higher home environment scores were directly associated with greater likelihood of employment in young adulthood.

Table 5
Moderated Mediation: Employment

	Coefficient	*SE*	95% CI
Effect on employment			
Education treatment	.30	0.24	−0.24, 0.72
Risk	−.23*	0.12	−0.45, −0.01
Home environment	.25*	0.12	−0.01, 0.46
Treatment × Home	.07	0.24	−0.42, 0.51
Effect on home environment			
Risk	−.33***	0.06	−0.49, −0.23
Indirect effect			
Indirect effect	−.08	0.04	−0.17, −0.00

**p* < .05. ****p* < .001.

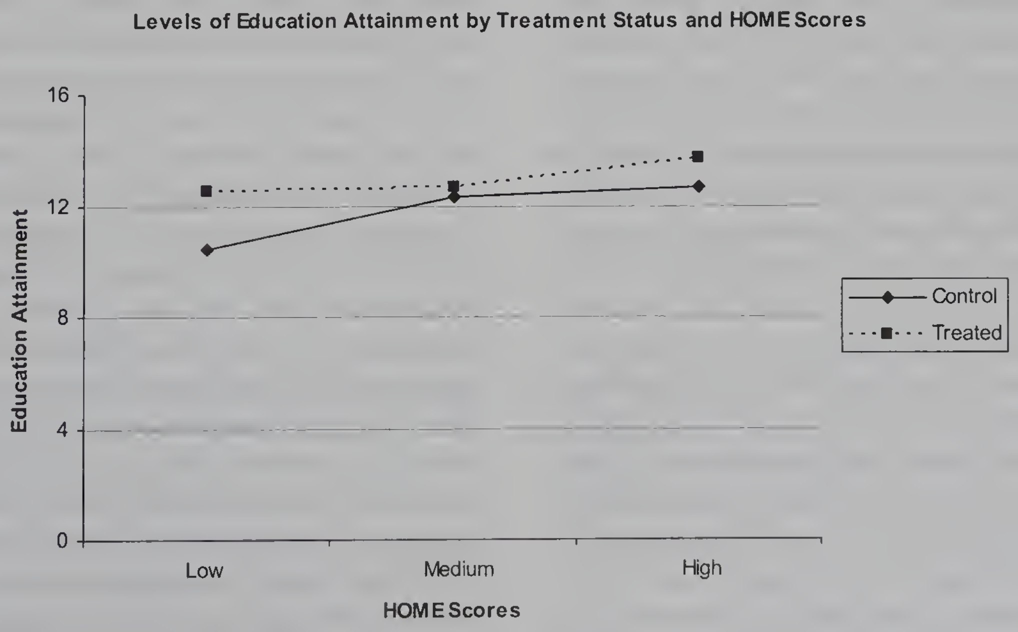


Figure 3. Years of education attainment by treatment status and Home Observation for Measurement of the Environment (HOME) scores.

Teen parenthood. Given the simultaneous estimation of effects, there was no evidence that the likelihood of having a child while a teen was related to risk level, the home environment, early education, nor an interactive effect of early education and the home environment.

Discussion

The first goal of the study was to examine the very long term effects of early cumulative risk over and above any effects of early intervention and to learn whether early intervention and early risk interacted when predicting the selected outcomes. As previously reported (Campbell et al., 2008), for the combined Abecedarian and CARE samples, early educational intervention was significantly associated with educational attainment in general, attending a 4-year college or university specifically, and obtaining skilled employment. The present analyses showed, in addition, that a prospective measure of cumulative risk summed across the first 5 years of life negatively predicted overall educational attainment, high school graduation specifically, and being employed as a young adult. Unexpectedly, early intervention was not found to moderate associations between early risk and young adult outcomes within this sample.

The second goal of this study was to examine a moderated mediation hypothesis: whether the quality of the early home mediated the found effects of early cumulative risk on later outcomes and whether early educational intervention moderated the effects of the early home environment on these outcomes. The results suggest that the home environment did mediate the effects of early risk for high school graduation but not for being employed or teen parenthood. Support for the moderated mediation hypothesis was found only for education attainment in general: The early home environment appeared to mediate the association between early risk and this outcome for the control group but not the treated group.

This study found associations between early risk and young adult accomplishments, a longer time span than is typical in the literature. The finding that early risk significantly predicted overall educational attainment and high school graduation specifically confirms as well as extends earlier work showing increased early cumulative risk to be associated with poor concurrent and shorter term academic outcomes (Brody et al., 2003; Forehand et al., 1998; Gutman et al., 2002; Gutman et al., 2003; Prelow & Loukas, 2003). The finding that higher early risk was associated with an increased likelihood of teen parenthood adds to previous work showing how individual risk factors are associated with teen parenthood (Cairns & Cairns, 1994; Haveman, Wolfe, & Wilson, 1997; Ludtke, 1997; Miller-Johnson, Winn, Coie, Malone, & Lochman, 2004). The current finding is modified by the fact that, within this sample, teen parenthood was no longer significantly predicted by risk when models simultaneously examined early risk, the early home environment, and early treatment.

The present findings suggest that when early cumulative risk and intensive early education in a child-care setting are considered simultaneously, higher level accomplishments in young adulthood were affected by early educational intervention whereas more basic-level accomplishments were associated with early risk. High school graduation, being employed as a young adult, and teen parenthood were all predicted by early cumulative risk irrespective of early intervention, whereas going to college and having a skilled-level job in young adulthood were predicted by early treatment, irrespective of early risk. Perhaps early risk added little to the prediction of going to college because it was at the earlier stage of completing high school that risk did the most harm, precluding the possibility of going to college. This is consistent with the report by Teachman, Paasch, Day, and Carver (1997) that family poverty appears to exert its greatest impact on high school graduation rather than on college attendance. Thus, within a high-risk sample, those experiencing the higher levels of risk find it harder to achieve some of the basic accomplishments of young adulthood, such as graduating from high school or getting any job. Early intervention, however, may provide the boost needed for higher levels of success, such as attending college or obtaining skilled employment.

A key finding from the present analyses is that treatment moderated the mediation of risk through the quality of the home environment. This effect emerged when predicting to the linear measure of young adult education attainment. For children in the control group, early cumulative risk was associated with a poorer quality home environment, which in turn was associated with lower levels of education attainment in young adulthood. This was consistent with our hypotheses. In contrast, our analyses found no evidence of such mediation for the treated group. Having the 5 years of educational intervention in a high-quality child-care setting appeared to be protective; that is, it buffered treated children against the long-term effects of a poor-quality early home environment on later educational attainment. This finding is consistent with other work suggesting that early intervention may moderate the effects of the early home environment (Bradley et al., 2001). Previous research involving the effects of child care on shorter term outcomes with high-risk samples has found evidence that child-care quality serves as a protective factor for some outcomes but not others. Burchinal et al. (2006) in their study of African American children in kindergarten through third grade found quality child care to be protective for math scores and behavior problems, but not for reading. Thus, different protective factors may be important for different outcomes (Gutman et al., 2003). The results here suggest that the effect for treatment was larger for those from poor-quality home environments than for those from higher quality early home environments. This implies that high-quality child care may be more effective for children whose parents are less responsive and provide less stimulating home environments than for other children.

Although evidence was found that early home environment mediated the effects of risk for educational outcomes, no evidence of such mediation was found for being employed or teen parenthood. The HOME, the measure used to assess the quality of the early home environment, has previously been shown to be associated with cognitive development and academic outcomes (Bradley, Caldwell, & Rock, 1988). Given this, its association with later educational outcomes in the present sample is consistent with the literature. One mechanism to explain the significant mediation findings is suggested by the work of Brody et al. (2003), who studied young African American adolescents growing up in single-mother homes in rural settings. These investigators found that cumulative risk was associated with parenting practices, which were in turn associated with youth self-regulation. Self-regulation was in turn associated with academic achievement. Heckman, Stixrud, and Urzua (2006) have suggested that noncognitive factors, as well as cognitive factors, strongly predict educational attainment. The environment provided by one's parents, combined with inherited ability and early intervention all play a role (Heckman et al., 2006). However, in the present study, the links between the early home environment and later educational attainments appear to be different from those that influence being employed and adolescent parenthood. The decision to become sexually active is complex (Michels, Kropp, Eyre, & Halpern-Felsher, 2005). Concurrent peer relationships and other adolescent circumstances undoubtedly play crucial roles in decisions made about parenthood when an adolescent is faced with such a life-altering experience. Young adult employment is subject to concurrent employment opportunities as well as personal characteristics influencing one's desire to seek work. Thus, the effects of early cumulative risk on employment and teen parenthood outcomes need further study; the links may be direct, as found here, or mediated through other early or concurrent factors not assessed in this study.

The hypothesis that increased early risk would be positively associated with illegal activity was not confirmed within this sample. In the present study, crime was represented by the young adults' self-reported convictions for misdemeanors and felonies and use of illegal drugs. A large body of research links risk and concurrent behavior problems. For example, using an index of cumulative risk, Furstenberg, Cook, Eccles, Elder, and Sameroff (1999) found that urban adolescents with 8 or more risk indicators (out of 10 possible) showed a 40% rate of behavior problems, compared to only 7% of those with 2 risk factors. However, these authors were examining current rates of problem behavior compared with current risk circumstances. Perhaps concurrent risk is more predictive of these outcomes than cumulative risk experienced much earlier in the life span.

No evidence was found that early educational intervention and early risk directly interacted to influence young adult outcomes. This finding was unexpected. It is noteworthy in this regard that the present sample consisted entirely of individuals who were drawn from high-risk backgrounds. Whether children exposed to more risk gain more from quality child care is a research question pertinent to policies regarding preferential entrance into public child-care programs, such as Head Start and prekindergarten, many of which use a high-risk index to determine who is recruited. Prior observational studies have found quality child care to be a stronger predictor of positive outcomes for children exposed to more social risk during early childhood (Burchinal et al., 2006; Burchinal, Peisner-Feinberg, Bryant, & Clifford, 2000; Caughy, DiPietro, & Strobino, 1994; Hubbs-Tait et al., 2002; Peisner-Feinberg & Burchinal, 1997; Peisner-Feinberg et al., 2001; Schliecker, White, & Jacobs, 1991; Vandell, 2004). This sample, however, consisted of individuals all considered high risk due to poverty and most considered at high risk due to minority ethnic status. This homogeneity within the present sample could have precluded detection of the expected interaction between risk and treatment.

Some limitations should be considered in interpreting the findings of the study. First, the sample was limited to individuals born into low-income families almost all of whom were African American, and thus the findings generalize to that demographic group and may apply to other groups in unknown ways. As noted earlier, being minority ethnic status is a variable that in and of itself has been found to be a risk factor. In addition to being more likely to be of low income (Denavas-Walt, Proctor, & Lee, 2006) and to experience other stressors that European Americans (Evans, 2004), discrimination and racism experienced by ethnic minority families may also negatively influence development (Murry, Brown, Brody, Cutrona, & Simons, 2001; Prelow, Danoff-Burg, Swenson, & Pulgiano, 2004). That is, as Murry, Brown, et al. (2001) found, although a lack of income and other resources is a significant factor in families' stress and functioning, ''simply being Black in America'' (p. 917) also played a critical yet sometimes unacknowledged role affecting maternal psychological functioning, thus influencing the quality of relationships and interactions in the home environment.

A second caveat is that the sample size was small, which may have reduced the power to detect smaller relationships and interaction effects (McClellan & Judd, 1993). This may have been especially problematic when considering the data on illegal activity. The total sample size was relatively small, and the proportion of young adults who self-reported lawbreaking was smaller still. Admission of a misdemeanor occurred in only 17% of the sample, and only 10% admitted a felony. Power may also have been limited by the fact that most of the outcomes considered here were dichotomous rather than continuous variables.

Another limitation of this work, one generally shared by the literature in this area, is that some of the risk factors that may have contributed to outcomes were not assessed and thus were not included in the cumulative risk index. Within the literature on this topic, wide variety exists concerning this issue. For example, Brody et al. (2003) included 7 possible factors, Sameroff et al. (1993) had 10, and Evans (2003) listed 9. This diversity among risk factors makes it difficult to compare findings across studies. An alternate approach to studying the question of risk might be to combine risk variables empirically using factor analyses as was performed by Deater-Decker et al. (1998). Burchinal, Roberts, et al. (2000) compared three approaches to analyzing risk—testing individual risk factors, creating risk factor scores using factor analyses, and calculating a cumulative risk index—and found that similar, but not identical, conclusions could be drawn using each approach. Given sufficient sample sizes, considering domains of risk, rather than simply summing indicators, may be more appropriate for different outcomes. However, longitudinal studies often consist of small sample sizes, for which a cumulative risk index may be most appropriate (Burchinal, Roberts, et al., 2000).

Similarly, the models in the current analyses did not include protective factors (beyond early educational intervention) that may have been related to outcomes, nor did they include how the adolescents and young adults interpreted the early risk factors. Future research could include these variables to gain a better understanding of the effects of early risk as well as pathways to competence in the context of adversity (Masten et al., 1999).

Another caveat concerns generalizations that can be made about the effects of child care. Although the treatment took place in a full-time child-care setting, the educational intervention was not typical of the early child-care experience for most young children. Moreover, although the treatment and control groups differed by the fact that one received the systematic early treatment and the other did not, many children in the control group did experience out-of-home care before the age of 5. Different types were used by this group including relative care, family day care homes, and state-licensed center care, as determined by family circumstances. Thus, the group comparisons are between those who did and did not receive early educational intervention in a child-care setting, not between those who did and did not experience out-of-home child care.

Despite the limitations, several strengths of this study increase confidence in the findings. These strengths include the study's longitudinal nature, which provided prospective data on early risk, and the very long span of time covered by the data collection, from birth to young adulthood. This permitted early risk to be summed across early childhood, capturing the natural variation that occurred within individual lives as experienced at the time. As noted by Gerard and Buehler (2004), one of the strengths of cumulative risk models is ''their potential to capture the natural covariation of risk factors'' (p. 1833). In addition, both the Abecedarian and CARE studies had low rates of attrition, increasing investigator confidence in long-term outcomes. Confidence in the conclusions pertaining to early educational intervention is increased by the fact that both Abecedarian and CARE were randomized studies. This degree of experimental control allows for the interpretation of early treatment's contribution to the outcomes with more confidence than would be possible in a naturalistic study where self-selection into treatment could have biased the findings.

These findings have implications for public policymakers considering how to allocate limited resources. Given that the effects of early cumulative risk can be very long-lasting and impact important basic young adult outcomes, resources are needed to help protect high-risk children from the effects of such multiple stressors. Increased early risk may have a particularly negative effect on some of the basic-level accomplishments society requires for minimal self-sufficiency. Early intervention, however, moderated the effects of less optimal home environments on educational attainment and was promotive for the important outcomes of college attendance and obtaining skilled employment in this high-risk sample. These findings affirm the allocation of resources to provide early childhood programs for high-risk children.

References

Aquilino, W. S. (1996). The life course of children born to unmarried mothers: Childhood living arrangements and young adult outcomes. *Journal of Marriage and the Family*, *58*, 293–310.

Arnett, J. J. (2000). Emerging adulthood: A theory of development from the late teens through the twenties. *American Psychologist*, *55*, 469–479.

Barocas, R., Seifer, R., Sameroff, A. J., Andrews, T. A., Croft, R. T., & Otrow, E. (1991). Social and interpersonal determinants of developmental risk. *Developmental Psychology*, *27*, 479–488.

Berrueta-Clement, J. R., Schweinhart, L. J., Barnett, W. S., Epstein, A. S., & Weikart, D. P. (1984). *Changed lives: The effects of the Perry Preschool Program on youths through age 19*. Ypsilanti, MI: High/Scope Press.

Bradley, R. H. (1992). *A factor analytic study of the HOME inventory in Black, White and Hispanic Americans*. Paper presented at the International Conference on Children at Risk, Santa Fe, NM.

Bradley, R. H., & Caldwell, B. M. (1979). Home Observation for Measurement of the Environment: A revision of the preschool scale. *American Journal of Mental Deficiency*, *84*, 235–244.

Bradley, R. H., Caldwell, B. M., & Rock, S. L. (1988). Home environment and school performance: A ten-year follow-up and examination of three models of environmental action. *Child Development*, *59*, 852–867.

Bradley, R. H., Caldwell, B. M., Rock, S. L., Hamrick, H. M., & Harris, P. (1988). Home Observation for Measurement of the Environment: Development of a home inventory for use with families having children 6 to 10 years old. *Contemporary Educational Psychology*, *13*, 58–71.

Bradley, R. H., Corwyn, R. F., Burchinal, M., McAdoo, H. P., & Garcia-Coll, C. (2001). The home environments of children in the United States: Part 2. Relations with behavioral development from birth through age 13. *Child Development*, *72*, 1868–1886.

Brody, G. H., & Flor, D. L. (1998). Maternal resources, parenting practices, and child competence in rural, single-parent African American families. *Child Development*, *69*, 803–816.

Brody, G. H., Kim, S., & Murry, V. M. (2003). Longitudinal links between contextual risks, parenting, and youth outcomes in rural African American families. *Journal of Black Psychology*, *29*, 359–377.

Bronfenbrenner, U., & Morris, P. A. (1998). The ecology of developmental processes. In W. Damon (Series Ed.) & R. M. Lerner (Vol. Ed.), *Handbook of child psychology: Vol. 1. Theoretical models of human development* (5th ed., pp. 993–1028). New York: Wiley.

Bryant, D. M., Burchinal, M., Lau, L. B., & Sparling, J. J. (1994). Family and classroom correlates of Head Start children's developmental outcomes. *Early Childhood Research Quarterly*, *9*, 289–309.

Burchinal, M. R., Campbell, F. A., Bryant, D. M., Wasik, B. H., & Ramey, C. T. (1997). Early intervention and mediating processes in cognitive performance of children of low-income African American families. *Child Development*, *68*, 935–954.

Burchinal, M., Peisner-Feinberg, E., Bryant, D., & Clifford, R. (2000). Children's social and cognitive development and child-care quality: Testing for differential associations related to poverty, gender, or ethnicity. *Applied Developmental Science*, *4*, 149–165.

Burchinal, M. R., Roberts, J. E., Hooper, S., & Zeisel, S. A. (2000). Cumulative risk and early cognitive development: A comparison of statistical risk models. *Developmental Psychology*, *36*, 793–807.

Burchinal, M., Roberts, J. E., Zeisel, S. A., Hennon, E. A., & Hooper, S. (2006). Social risk and protective child, parenting, and child care factors in early elementary school years. *Parenting: Science and Practice*, *6*, 79–113.

Cairns, R. B., & Cairns, B. D. (1994). *Lifelines and risks: Pathways of youth in our time*. Cambridge, UK: Cambridge University Press.

Caldwell, B. M., & Bradley, R. H. (1984). *Home observation for measurement of the environment*. Little Rock: University of Arkansas.

Campbell, F. A., Pungello, E. P., Miller-Johnson, S., Burchinal, M., & Ramey, C. T. (2001). The development of cognitive and academic abilities: Growth curves from an early childhood educational experiment. *Developmental Psychology, 37*, 231–242.

Campbell, F. A., Ramey, C. T., Pungello, E. P., Sparling, J., & Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian Project. *Applied Developmental Science, 6*, 42–57.

Campbell, F. A., Wasik, B. H., Pungello, E. P., Burchinal, M. R., Kainz, K., Barbarin, O., et al. (2008). Young adult outcomes from the Abecedarian and CARE early childhood educational interventions. *Early Childhood Research Quarterly, 23*, 452–466.

Caughy, M. O., DiPietro, J. A., & Strobino, D. M. (1994). Day-care participation as a protective factor in the cognitive development of low-income children. *Child Development, 65*, 457–471.

Chase-Lansdale, P. L., Gordon, R. A., Brooks-Gunn, J., & Klebanov, P. K. (1997). Neighborhood and family influences on the intellectual and behavioral competence of preschool and early school-age children. In J. Brooks-Gunn, G. J. Duncan, & J. L. Aber (Eds.), *Neighborhood poverty: Vol. 1. Context and consequences for children* (pp. 79–118). New York: Russell Sage Foundation.

Conger, R. D., & Elder, G. H. (1994). Families in troubled times: The Iowa Youth and Families Project. In R. D. Conger & G. H. Elder Jr. (Eds.), *Troubled times: Adapting to change in rural America* (pp. 3–19). New York: Aldine.

Deater-Decker, K., Dodge, K. A., Bates, J. E., & Pettit, G. S. (1998). Multiple risk factors in the development of externalizing behavior problems: Group and individual differences. *Development and Psychopathology, 10*, 469–493.

Denavas-Walt, C., Proctor, B. D., & Lee, C. H. (2006). *Income, poverty, and health insurance coverage in the United States* (U.S. Census Bureau, Current Population Reports, P60-231). Washington, DC: U.S. Government Printing Office.

Dubow, E. F., & Luster, T. (1990). Adjustment of children born to teenage mothers: The contributions of risk and protective factors. *Journal of Marriage and the Family, 52*, 393–404.

Ensiminger, M. E., & Slusarcick, A. L. (1992). Paths to high school graduation or dropout: A longitudinal study of a first-grade cohort. *Sociology of Education, 65*, 95–113.

Evans, G. W. (2003). A multimethodological analysis of cumulative risk and allostatic load among rural children. *Developmental Psychology, 39*, 924–933.

Evans, G. W. (2004). The environment of childhood poverty. *American Psychologist, 59*, 77–92.

Forehand, R., Biggar, H., & Kochick, B. A. (1998). Cumulative risk across family stressors: Short- and long-term effects for adolescents. *Journal of Abnormal Child Psychology, 26*, 119–128.

Furstenberg, F. F., Brooks-Gunn, J., & Chase-Lansdale, L. (1989). Teen-aged pregnancy and childbearing. *American Psychologist, 44*, 313–320.

Furstenberg, F., Cook, T., Eccles, J., Elder, G., & Sameroff, A. (1999). *Managing to make it: Urban families in adolescent success*. Chicago: University of Chicago Press.

Garmezy, N., Masten, A. S., & Tellegen, A. (1984). The study of stress and competence in children: A building block for developmental psychopathology. *Child Development, 55*, 97–111.

Gerard, J. M., & Buehler, C. (2004). Cumulative environmental risk and youth maladjustment: The role of youth attributes. *Child Development, 75*, 1832–1849.

Gest, S. D., Mahoney, J. L., & Cairns, R. B. (1999). A developmental approach to prevention research: Early adolescence configurations associated with teenage parenthood. *American Journal of Community Psychology, 27*, 543–565.

Guo, G., & Harris, K. M. (2000). The mechanisms mediating the effects of poverty on children's intellectual development. *Demography, 37*, 431–447.

Guo, G., & VanWey, L. K. (1999). Sibship size and intellectual development: Is the relationship causal? *American Sociological Review, 64*, 169–187.

Gutman, L. M., Sameroff, A. J., & Cole, R. (2003). Academic trajectories from first to twelfth grades: Growth curves according to multiple risk and early child factors. *Developmental Psychology, 39*, 777–790.

Gutman, L. M., Sameroff, A. J., & Eccles, J. S. (2002). The academic achievement of African American students during early adolescence: An examination of multiple risk, promotive, and protective factors. *American Journal of Community Psychology, 30*, 367–399.

Haurin, R. J. (1992). Patterns of childhood residence and the relationship to young adult outcomes. *Journal of Marriage and the Family, 54*, 846–860.

Haveman, R., Wolfe, B., & Wilson, K. (1997). Childhood poverty and adolescent schooling and fertility outcomes: Reduced-form and structural estimates. In G. Duncan & J. Brooks-Gunn (Eds.), *Consequences of growing up poor* (pp. 419–460). New York: Russell Sage Foundation.

Heckman, J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science, 312*, 1900–1902.

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics, 24*, 411–482.

Hollingshead, A. B. (n.d.). *Four Factor Index of Social Status* (Working paper). New Haven, CT: Yale University, Department of Sociology.

Hooper, S. R., Burchinal, M. R., Roberts, J. E., Zeisel, S., & Neebe, E. (1998). Social and family risk factors for infant development at one year: An application of the cumulative risk model. *Journal of Applied Developmental Psychology, 19*, 85–96.

Hubbs-Tait, L., Culp, A. M., Huey, E., Culp, R. E., Starost, H. J., & Hare, C. (2002). Relation of Head Start attendance to children's cognitive and social outcomes:

Moderation by family risk. *Early Childhood Research Quarterly, 17*, 539–558.

Jaffee, S. (2002). Pathways to adversity in young adulthood among early childbearers. *Journal of Family Psychology, 16*, 38–49.

Jimerson, S., Egeland, B., Sroufe, A., & Carlson, B. (2000). A prospective longitudinal study of high school dropouts: Examining multiple predictors across development. *Journal of School Psychology, 38*, 535–549.

Jones, D. J., Forehand, R., Brody, G., & Armistead, L. (2002). Psychosocial adjustment of African American children in single-mother families: A test of three risk models. *Journal of Marriage and Family, 64*, 105–115.

Kolbe, L. J. (1990). An epidemiological surveillance system to monitor the prevalence of youth behaviors that most affect health. *Health Education, 21*, 44–48.

Krishnakumar, A., & Black, M. M. (2002). Longitudinal predictors of competence among African American children: The role of distal and proximal risk factors. *Journal of Applied Developmental Psychology, 23*, 237–266.

Liaw, F., & Brooks-Gunn, J. (1994). Cumulative familial risks and low-birthweight children's cognitive and behavioral development. *Journal of Clinical and Child Psychology, 23*, 360–372.

Linver, M., Brooks-Gunn, J., & Kohen, D. (2002). Family processes as pathways from income to young children's development. *Developmental Psychology, 38*, 719–734.

Ludtke, L. (1997). *On our own: Unmarried motherhood in America*. New York: Random House.

Luster, T., & McAdoo, H. P. (1994). Factors related to the achievement and adjustment of young African American children. *Child Development, 65*, 1080–1094.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. Mahwah, NJ: Erlbaum.

Masten, A. S., Best, K. M., & Garmezy, N. (1990). Resilience and development: Contributions from the study of children who overcome adversity. *Development and Psychopathology, 2*, 425–444.

Masten, A. S., Coatsworth, D. A., Neemann, J., Gest, S. D., Tellegen, A., & Garmezy, N. (1995). The structure and coherence of competence from childhood to adolescence. *Child Development, 66*, 1635–1659.

Masten, A. S., Hubbard, J., Gest, S. D., Tellegen, A., Garmezy, N., & Ramirez, M. (1999). Adversity, resources and resilience: Pathways to competence from childhood to late adolescence. *Development and Psychopathology, 11*, 143–169.

McClellan, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin, 114*, 376–390.

McGinness, G. D., & Ramey, C. T. (1981). Developing sociolinguistic competence in children. *Canadian Journal of Early Childhood Education, 1*, 22–43.

McLanahan, S. (1997). Parent absence or poverty: Which matters worse? In G. J. Duncan & J. Brooks-Gunn (Eds.), *Consequences of growing up poor* (pp. 35–48). New York: Russell Sage Foundation.

Michels, T. M., Kropp, R. Y., Eyre, S. L., & Halpern-Felsher, B. L. (2005). Initiating sexual experiences: How do young adolescents make decisions regarding early sexual activity? *Journal of Research on Adolescence, 15*, 583–607.

Miller-Johnson, S., Winn, D.-M. C., Coie, J. D., Malone, P. S., & Lochman, J. (2004). Risk factors for adolescent pregnancy reports among African American males. *Journal of Research on Adolescence, 14*, 471–495.

Murry, V. M., Brown, P. A., Brody, G. H., Cutrona, C. E., & Simons, R. L. (2001). Racial discrimination as a moderator of the links among stress, maternal psychological functioning and family relationships. *Journal of Marriage and the Family, 63*, 915–926.

Murry, V. M., Bynum, M. S., Brody, G. H., Willert, A., & Stephens, D. (2001). African American single mothers and children in context: A review of studies on risk and resilience. *Clinical Child and Family Psychology Review, 4*, 133–155.

Muthén, L. K., & Muthén, B. O. (1998-2007). Mplus users guide (5th ed.). Los Angeles: Author.

National Institute of Child Health and Human Development Early Child Care Research Network. (2005). Duration and developmental timing of poverty and children's cognitive and social development from birth through third grade. *Child Development, 74*, 795–810.

Peisner-Feinberg, E., & Burchinal, M. (1997). Relations between preschool children's child care experiences and concurrent development: The cost quality, and outcomes study. *Merrill-Palmer Quarterly, 43*, 451–477.

Peisner-Feinberg, E., Burchinal, M., Clifford, R., Culkin, M., Howes, C., Kagan, S., et al. (2001). The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second grade. *Child Development, 72*, 1534–1553.

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research, 42*, 185–227.

Prelow, H. M., Danoff-Burg, S., Swenson, R. R., & Pulgiano, D. (2004). The impact of ecological risk and perceived discrimination on the psychological adjustment of African American and European American youth. *Journal of Community Psychology, 32*, 375–389.

Prelow, H. M., & Loukas, A. (2003). The role of resource, protective, and risk factors on academic achievement-related outcomes of economically disadvantaged Latino youth. *Journal of Community Psychology, 31*, 513–529.

Pungello, E. P., Kupersmidt, J. B., Burchinal, M. R., & Patterson, C. J. (1996). Environmental risk factors and children's achievement from middle childhood to early adolescence. *Developmental Psychology, 32*, 755–767.

Ramey, C. T., & Campbell, F. A. (1984). Preventive education for high-risk children: Cognitive consequences of the Carolina Abecedarian Project. *American Journal of Mental Deficiency, 88*, 515–523.

Ramey, C. T., & Smith, B. (1977). Assessing the intellectual consequences of early intervention with high-risk infants. *American Journal of Mental Deficiency, 81*, 318–324.

Rutter, M. (1979). Protective factors in children's responses to stress and disadvantage. In M. W. Kent & J. E. Rolf (Eds.), *Primary prevention of psychopathology: Vol. 3. Social competence in children* (pp. 49–74). Hanover, NH: University Press of New England.

Sameroff, A. J., & MacKenzie, M. J. (2003). Research strategies for capturing transactional models of development: The limits of the possible. *Development and Psychopathology, 15*, 613–640.

Sameroff, A. J., Seifer, R., Baldwin, A., & Baldwin, C. (1993). Stability of intelligence from preschool to adolescence: The influence of social and family risk factors. *Child Development, 64*, 80–97.

Sameroff, A. J., Seifer, R., Barocas, R., Zax, M., & Greenspan, S. (1987). Intelligence quotient scores of 4-year-old children: Social environmental risk factors. *Pediatrics, 79*, 343–350.

Schliecker, E., White, D., & Jacobs, E. (1991). The role of day care quality in the prediction of children's vocabulary. *Canadian Journal of Behavioral Science, 23*, 12–24.

Schoon, I. E., Bynner, J., Joshi, H., Parsons, S., Wiggins, R. D., & Sacker, A. (2002). The influence of context, timing, and duration of risk experiences for the passage from childhood to mid-adulthood. *Child Development, 73*, 1486–1504.

Schweinhart, L. J., Barnes, H. V., & Weikart, D. P. (1993). Significant benefits: The High/Scope Perry Preschool Study through age 27. *Monographs of the High/Scope Educational Research Foundation, 10*. Ypsilanti, MI: High/Scope.

Shaw, D. S., Winslow, E. B., Owens, E. B., & Hood, N. (1998). Young children's adjustment to chronic family adversity: A longitudinal study of low-income families. *Journal of the American Academy of Child and Adolescent Psychiatry, 37*, 545–553.

Sparling, J. J. (1989). Narrow- and broad-spectrum curricula, two necessary parts of the special child's program. *Infants and Young Children, 1*, 1–8.

Sparling, J. J., & Lewis, I. (1978). *Learningames for the first three years: A guide to parent-child play*. New York: Walker.

Sparling, J. J., & Lewis, I. (1984). *Learningames for threes and fours: A guide to adult and child play*. New York: Walker.

Sugawara, A. I. (1991). Selected child factors mediating the impact of maternal absence on children's behaviors and development. *Early Child Development and Care, 72*, 1–22.

Teachman, K. M., Paasch, R., Day, R. D., & Carver, K. P. (1997). Poverty during adolescence and subsequent educational attainment. In G. J. Duncan & J. Brooks-Gunn (Eds.), *Consequences of growing up poor* (pp. 382–418). New York: Russell Sage Foundation.

Tein, J. Y., Sandler, I. N., MacKinnon, D. P., & Wolchik, S. A. (2004). How did it work? Who did it work for? Mediation in the context of a moderated prevention effect for children of divorce. *Journal of Consulting and Clinical Psychology, 72*, 617–624.

Trentacosta, C. J., Hyde, L. W., Shaw, D. S., Dishion, T. J., Gardner, F., & Wilson, M. (2008). The relations among cumulative risk, parenting, and behavior problems during early childhhod. *Journal of Child Psychology and Psychiatry, 49*, 1211–1219.

Vandell, D. L. (2004). Early child care: The known and the unknown. *Merrill-Palmer Quarterly, 50*, 387–414.

Wasik, B. H., Ramey, C. T., Bryant, D., & Sparling, J. (1990). A longitudinal study of two early intervention strategies: Project CARE. *Child Development, 61*, 1682–1696.

Williams, S., Anderson, J., McGee, R., & Silva, P. A. (1990). Risk factors for behavioral and emotional disorder in preadolescent children. *Journal of the American Academy of Child and Adolescent Psychiatry, 29*, 413–419.

Xie, H., Cairns, B. D., & Cairns, R. B. (2001). Predicting teen motherhood and teen fatherhood: Individual characteristics and peer affiliations. *Social Development, 10*, 488–511.

Zill, N., Morrison, D. R., & Coiro, M. J. (1993). Long-term effects of parental divorce on parent-child relationships, adjustment, and achievement in young adults. *Journal of Family Psychology, 7*, 91–103.

Child Development, January/February 2010, Volume 81, Number 1, Pages 427–429

Manuscripts Accepted for Publication

Lying and Truth-Telling in Children: From Concept to Action, Fen Xu, Xuehua Bao, Victoria Talwar, and Kang Lee.

Preschoolers' Implicit and Explicit False-Belief Understanding: What Are the Relations With Complex Syntax? Jason Low.

Mother–Infant Interaction During the First 3 Months: The Emergence of Culture-Specific Contingency Patterns, Joscha Kärtner, Heidi Keller, and Relindis Yovsi.

Not Doing What You Are Told: Early Perseverative Errors in Updating Mental Representations via Language, Patricia A. Ganea and Paul L. Harris.

The Rise in Cortisol in Family Day Care: Associations With Aspects of Care Quality, Child Behavior, and Child Sex, Megan Gunnar, Erin Kryzer, Mark J. Van Ryzin, and Deborah A. Phillips.

"My Way or Mom's Way?": The Bilingual and Bicultural Self in Hong Kong Chinese Children and Adolescents, Qi Wang, Yi Shao, and Yexin Jessica Li.

The Significance of Insecure Attachment and Disorganization in the Development of Children's Externalizing Behavior: A Meta-Analytic Study, Pasco Fearon, Marian Bakermans-Kranenburg, Marinus Van IJzendoorn, Anne-Marie Lapsley, and Glenn Roisman.

Children's Sensitivity to the Conventionality of Sources, Gil Diesendruck, Nurit Carmel, and Lori Markson.

When Do Children Exhibit a Yes Bias? Mako Okanda and Shoji Itakura.

Children Monitor Individuals' Expertise for Word Learning, David Sobel and Kathleen Corriveau.

Developmental Patterns in Decision-Making Autonomy Across Middle Childhood and Adolescence: European American Parents' Perspectives, Laura Wray-Lake, Ann C. Crouter, and Susan M. McHale.

Bridging Theory of Mind and the Personal Domain: Children's Reasoning About Resistance to Parental Control, Kristin Lagattuta, Larry Nucci, and Sandra Leanne Bosacki.

The Dynamics of Free Recall and Their Relation to Rehearsal Between 8 and 10 Years of Age, Martin Lehmann and Marcus Hasselhorn.

The Tuning of Human Neonates' Preference for Speech, Athena Vouloumanos, Marc D. Hauser, Janet F. Werker, and Alia Martin.

Early Gesture Predicts Language Delay in Children With Pre- or Perinatal Brain Lesions, Eve Sauer, Susan C. Levine, and Susan Goldin-Meadow.

Coercive and Prosocial Fathering, Antisocial Personality, and Growth in Children's Postdivorce Noncompliance, David S. DeGarmo.

Parent Discipline Practices in an International Sample: Associations With Child Behaviors and Moderation by Perceived Normativeness, Elizabeth Gershoff, Andrew Grogan-Kaylor, Jennifer E. Lansford, Chang Lei, Arnaldo Zelli, Kirby Deater-Deckard, and Kenneth A. Dodge.

The Complex Relation Between Bullying, Victimization, Acceptance, and Rejection: Giving Special Attention to Status, Affection, and Sex Differences, René Veenstra, Siegwart Lindenberg, Anke Munniksma, and Jan Kornelis Dijkstra.

Who Said You Could Wear My Sweater?: Adolescent Siblings' Conflicts and Associations With Relationship Quality, Nicole Campione-Barr and Judith G. Smetana.

Categorization in 3- and 4-Month-Old Infants: An Advantage of Words Over Tones, Alissa Ferry, Susan Hespos, and Sandra Waxman.

Family Socioeconomic Status and Consistent Environmental Stimulation in Early Childhood, Robert Crosnoe, Tama Leventhal, R. J. Wirth, Kim Pierce, and Robert Pianta.

Defining Best Practice in Care and Protection of Children in Crisis-Affected Settings: A Delphi Study, Alastair Ager, Lindsay Stark, Bree Akesson, and Neil Boothby.

Inconsistency With Prior Knowledge Triggers Children's Causal Explanatory Reasoning, Cristine H. Legare, Susan A. Gelman, and Henry M. Wellman.

Children's Tattling: The Reporting of Everyday Norm Violations in Preschool Settings, Gordon Ingram and Jesse Bering.

Conversation, Development, and Gender: A Study of Changes in Children's Concepts of Punishment, Patrick Leman and Marina Bjornberg.

Integrating Different Perspectives on Socialization Theory and Research: A Domain-Specific Approach, Joan Grusec and Maayan Davidov.

Children's Sleep and Adjustment Over Time: The Role of Socioeconomic Context, Mona El-Sheikh, Ryan J. Kelly, Joseph A. Buckhalt, and J. Benjamin Hinnant.

Antecedents and Correlates of the Popular-Aggressive Phenomenon in Elementary School, Philip Rodkin and Glenn I. Roisman.

The Effects of the 1999 Turkish Earthquake on Young Children: Analyzing Traumatized Children's Completion of Short Stories, Elif Celebi Oncu and Aysegul Metindogan Wise.

Individual and Peer Group Normative Beliefs About Relational Aggression: Developmental Changes and Prediction of Relational Aggression, Nicole E. Werner and Laura G. Hill.

Gender, Values, and Occupational Interests Among Children, Adolescents, and Adults, Erica Weisgram, Rebecca S. Bigler, and Lynn S. Liben.

Children's Understanding of Social-Cognitive and Social-Communicative Aspects of Discourse Irony, Eva Filippova and Janet Wilde Astington.

The Head Bone's Connected to the Neck Bone: When Do Toddlers Represent Their Own Body Topography? Celia A. Brownell, Sara R. Nichols, Margarita Svetlova, Stephanie Zerwas, and Geetha Ramani.

Connecting Cues: Overlapping Regularities Support Cue Discovery in Infancy, Sarah Sahni, Mark Seidenberg, and Jenny Saffran.

Mexican-Heritage Children's Attention and Learning From Interactions Directed to Others, Katie G. Silva, Maricela Correa-Chávez, and Barbara Rogoff.

The Development of Social Essentialism: The Case of Israeli Children's Inferences About Jews and Arabs, Dana Birnbaum, Inas Deeb, Gili Segall, Adar Ben-Eliyahu, and Gil Diesendruck.

Parent Involvement and Children's Academic and Social Development in Elementary School, Nermeen El Nokali, Heather J. Bachman, and Elizabeth Votruba-Drzal.

Do Effects of Early Child Care Extend to Age 15 Years? Deborah Vandell, Jay Belsky, Margaret Burchinal, Laurence Steinberg, Nathan Vandergrift, and NICHD Early Child Care Research Network.

I Like Me If You Like Me: On the Interpersonal Modulation and Regulation of Preadolescents' State Self-Esteem, Sander Thomaes, Albert Reijntjes, Bram Orobio de Castro, Brad J. Bushman, Astrid Poorthuis, and Michael J. Telch.

An Eye for Detail: An Event-Related Potential Study of the Rapid Processing of Fearful Facial Expressions in Children, Petra Vlamings, Lisa M. Jonkman, and Chantal Kemner.

Experience-Based and On-Line Categorization of Objects in Early Infancy, Marc Bornstein and Clay Mash.

On the Progression and Stability of Adolescent Identity Formation, Wim Meeus, Rens van de Schoot, Loes Keijsers, Seth J. Schwartz, and Susan Branje.

Dogs, Bogs, Labs, and Lads: What Phonemic Generalizations Indicate About the Nature of Children's Early Word-Form Representations, Erik Thiessen and Meagan N. Yee.

Investigating the Development of Core Cognitive Skills in Autism: A 3-Year Prospective Study, Elizabeth Pellicano.

Home Media and Children's Achievement and Behavior, Sandra Hofferth.

The Effects of Maltreatment and Neuroendocrine Regulation on Memory Performance, Dante Cicchetti, Fred A. Rogosch, Mark L. Howe, and Sheree L. Toth.

Interracial and Intraracial Contact, School-Level Diversity, and Change in Racial Identity Status Among African American Adolescents, Tiffany Yip, Eleanor K Seaton, and Robert M. Sellers.

Child Care and the Development of Behavior Problems Among Economically Disadvantaged Children in Middle Childhood, Elizabeth Votruba-

Drzal, Rebekah Coley, Carolina Maldonado Carreno, Christine Li-Grining, and P. Lindsay Chase-Lansdale.

Children's Classroom Engagement and School Readiness Gains in Prekindergarten, Nina C. Chien, Carollee Howes, Margaret Burchinal, Robert Pianta, Sharon Ritchie, Donna Bryant, Richard Clifford, Diane Early, and Oscar Barbarin.

Early Elementary School Adjustment of Maltreated Children in Foster Care: The Roles of Inhibitory Control and Caregiver Involvement, Katherine Pears, Philip A. Fisher, Jacqueline Bruce, Hyoun K. Kim, and Karen Yoerger.

Children's Understanding of Ordinary and Extraordinary Minds, Jonathan Lane, Henry Wellman, and E. Margaret Evans.

What Paradox? Referential Cues Allow For Infant Use of Phonetic Detail in Word Learning, Christopher Fennell and Sandra R. Waxman.

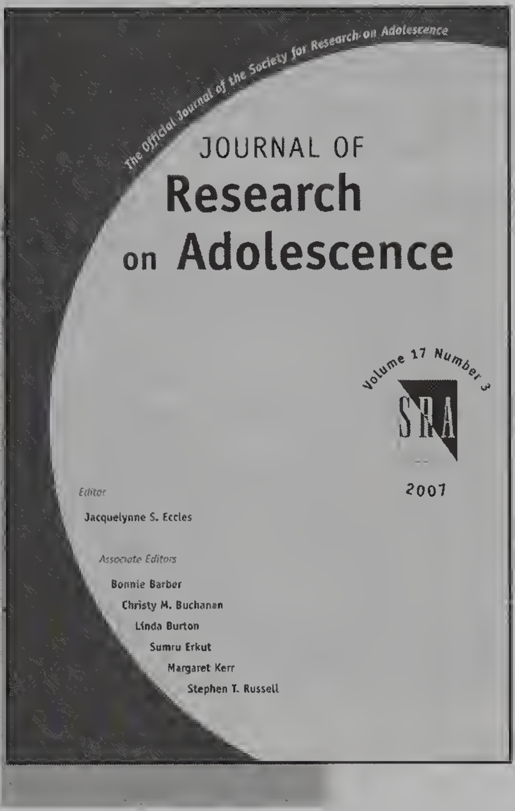
Contagious Yawning in Autistic and Typical Development, Molly Helt, Inge-Marie Eigsti, Peter J. Snyder, and Deborah A. Fein.

Tolerance of Practices by Muslim Actors: An Integrative Social-Developmental Perspective, Maike Gieling, Jochem Thijs, and Maykel Verkuyten.

Structure and Strategies in Children's Educational Television: The Roles of Program Type and Learning Strategies in Children's Learning, Deborah L. Linebarger and Jessica Taylor Piotrowski.

Bridging the Gap: Solving Spatial Means–Ends Relations in a Locomotor Task, Sarah Berger, Karen E. Adolph, and Alisan E. Kavookjian.

Sex Categorization Among Preschool Children: Increasing Utilization of Sexually Dimorphic Cues, Kerri Johnson, Leah Lurye Murphy, and Louis G. Tassinary.



The Official Journal of the Society for Research on Adolescence

JOURNAL OF
Research
on Adolescence

Volume 17 Number 3
SRA
2007

Editor
Jacquelynne S. Eccles

Associate Editors
Bonnie Barber
Christy M. Buchanan
Linda Burton
Sumru Erkut
Margaret Kerr
Stephen T. Russell

JOURNAL OF
Research
on Adolescence

Edited by
JACQUELYNNE S. ECCLES

The Official Journal of the Society for Research on Adolescence

Multidisciplinary and international in scope, the ***Journal of Research on Adolescence (JRA)*** significantly advances knowledge in the field of adolescent research. Employing a diverse array of methodologies, this compelling journal publishes original research and integrative reviews of the highest level of scholarship. Featured studies include both quantitative and qualitative methodologies applied to cognitive, physical, emotional, and social development and behavior. Articles pertinent to the variety of developmental patterns inherent throughout adolescence are featured, including cross-national and cross-cultural studies. Attention is given to normative patterns of behavior as well as individual differences rooted in personal or social and cultural factors.

WILEY-
BLACKWELL

For more information and to subscribe online visit
www.interscience.wiley.com/journal/jra

Notice to Contributors

Child Development publishes empirical, theoretical, review, applied, and policy articles reporting research on child development. Published by the interdisciplinary Society for Research in Child Development (SRCD), the journal welcomes relevant submissions from all disciplines. Further information is available at http://www.srcd.org/cd.html.

Types of Articles

Child Development considers manuscripts in formats described below. Inquiries concerning alternative formats should be addressed to the Editor prior to submission. All submissions are expected to be no more than 40 manuscript pages, including tables, references, and figures (but excluding appendices). Authors should provide a justification if the submission is substantially longer. Unless the editor finds that justification compelling, the submission will be returned to the author for shortening prior to editorial review.

Empirical articles comprise the major portion of the journal. To be accepted, empirical articles must be judged as being high in scientific quality, contributing to the empirical base of child development, and having important theoretical, practical, or interdisciplinary implications. Reports of multiple studies, methods, or settings are encouraged, but single-study reports are also considered. Empirical articles will thus vary considerably in length, but should be no longer than 40 manuscript pages; text and graphics should be as concise as material permits. All modes of empirical research are welcome.

Empirical reports are reserved for short cutting-edge empirical papers that are no longer than 4,000 words in length (including text, tables, footnotes, appendices), which advance research and knowledge in an area through noteworthy findings and/or new methods. For manuscripts that require longer descriptions of methods and results, authors should use the Empirical article format.

Reviews focus on past empirical and/or on conceptual and theoretical work. They are expected to synthesize or evaluate a topic or issue relevant to child development, should appeal to a broad audience, and may be followed by a small number of solicited commentaries.

Essays describe original concepts, methods, trends, applications, and theories; these may also be accompanied by solicited commentaries.

Child Development and ... are articles that provide readers with tutorials about some new concept or academic specialty pertinent to research in child development. These papers should review the major definitions, methods, and findings of the concept or specialty and discuss past or potential links to child development.

From another perspective is a format in which papers on a focal topic, written by different authors, are published simultaneously. Papers represent diverse perspectives (e.g., authors whose work represents different populations; different disciplines; different theories, methods, or analytic tools). In some cases, calls for submissions on particular topics will be disseminated through SRCD (via e-mail or SRCD publications), and submissions will undergo normal editorial review. In some cases, a submitted manuscript (e.g., an empirical article) may be selected as a lead article for this format, with invited commentaries providing additional perspectives. The editors also welcome suggestions from readers for topics for this format.

Manuscript Submission

Child Development invites for consideration manuscripts that are neither identical to nor substantially similar to work published or under review elsewhere. In the submission cover letter, please provide details about other published or submitted papers having substantial overlap (including data sets) with the new CD submission to enable editors to judge whether the new submission is sufficiently distinct from other work to warrant consideration. Please note if the paper is posted on a website, see http://www. srcd.org/web posting. html. Editors retain the right to reject manuscripts that do not meet established ethical standards for research or dissemination.

Manuscripts should be submitted electronically to the Child Development Online Submission Site at www.srcd.org /CDsubmit/ as a Word or WordPerfect file. Please also submit a cover letter that contains the name(s) of the author(s) and affiliation(s), and the street address, telephone, fax, and electronic mail address of the corresponding author. A corresponding author's submission to *Child Development* implies that all coauthors have agreed to the content and form of the manuscript and that the ethical standards of SRCD have been followed (see the *Child Development* website or pp. 283–284 of the 2000 SRCD Directory). Any financial interest or conflict of interest must be explained to the Editor in the cover letter. The corresponding author is responsible for informing all coauthors, in a timely manner, of manuscript submission, editorial decisions, reviews, and revisions. Please note that if your submission (or its revision) is accepted for publication, each author must submit a signed Exclusive License Form (http://www.srcd.org/Documents/Publications/CD/cdcopyright.pdf) before the manuscript can enter production. Planning ahead to meet this requirement is especially important for manuscripts with multiple authors.

The manuscript file should be formatted with double-spaced, 12-point type, and should include a single paragraph abstract of 100-120 words. Please follow all guidelines on format, style, and ethics provided in the Publication Manual (5th ed.) of the American Psychological Association. Figures included with initial submissions will not be returned. Therefore, please submit only electronic files or copies of figures. If color artwork is submitted, and the authors believe color art is necessary to the presentation of their work, the submissions letter should indicate that one or more authors or their institutions are prepared to pay the substantial costs associated with color art reproduction. Authors should keep a copy of all correspondence, files, and figures to guard against loss.

Manuscript Review

If you have any questions about your submission, please inquire at cdev@srcd.org or call (734) 926-0615. Each manuscript is handled by the Editor or an Associate Editor who consults with one or more Consulting Editors and/or ad hoc reviewers who have relevant expertise. To ensure blind review, cover sheets are removed before review; authors should avoid including any other information about identity or affiliation in submissions. Copies of the submission and associated correspondence are retained in the SRCD archives. For accepted manuscripts, authors are required to prepare a 300-500 layperson's summary for public dissemination purposes. Details are provided to authors as part of final processing.

There is no charge for publication in *Child Development* unless tabular or graphic materials exceed 10% of the total number of pages. Charges are also levied for changes in proofs other than correction of printer's errors. Any inquiries relating to charges or business matters (including reprint orders) should be addressed to Wiley-Blackwell, *Child Development*, Production Coordinator, 350 Main Street, Malden, MA 02148, (781) 388-8200.

Inquiries and suggestions regarding editorial policy may be addressed to: Dr. Jeffrey J. Lockman, Editor, *Child Development*, Department of Psychology, Tulane University, 2007 Percival Stern Hall, New Orleans, LA 70118-5698, lockman@tulane.edu. *Please do not send electronic or paper manuscripts to the above address.* Doing so will delay the processing of your manuscript.

Note to NIH Grantees

Pursuant to NIH mandate, Society through Wiley-Blackwell will post the accepted version of Contributions authored by NIH grantholders to PubMed Central upon acceptance. This accepted version will be made publicly available 12 months after publication. For further information, see www.wiley.com/go/nihmandate.

MONOGRAPHS OF THE SOCIETY FOR RESEARCH IN CHILD DEVELOPMENT

Recent

A DYNAMIC CASCADE MODEL OF THE DEVELOPMENT OF SUBSTANCE-USE ONSET
- *Kenneth A. Dodge, Patrick S. Malone, Jennifer E. Lansford, Shari Miller, Gregory S. Pettit, and John E. Bates* (Serial No. 294, 2009).

FLEXIBILITY IN EARLY VERB USE: EVIDENCE FROM A MULTIPLE-*N* DIARY STUDY
- *Letitia R. Naigles, Erika Hoff, and Donna Vear* (Serial No. 293, 2009).

MARITAL CONFLICT AND CHILDREN'S EXTERNALIZING BEHAVIOR: INTERACTIONS BETWEEN PARASYMPATHETIC AND SYMPATHETIC NERVOUS SYSTEM ACTIVITY
- *Mona El-Sheikh et al.* (Serial No. 292, 2009).

THE EFFECTS OF EARLY SOCIAL-EMOTIONAL AND RELATIONSHIP EXPERIENCE ON THE DEVELOPMENT OF YOUNG ORPHANAGE CHILDREN
- *The St. Petersburg–USA Orphanage Research Team* (Serial No. 291, 2008).

UNDERSTANDING MOTHER-ADOLESCENT CONFLICT DISCUSSIONS: CONCURRENT AND ACROSS-TIME PREDICTION FROM YOUTHS' DISPOSITIONS AND PARENTING
- *Nancy Eisenberg et al.* (Serial No. 290, 2008).

DEVELOPING OBJECT CONCEPTS IN INFANCY: AN ASSOCIATIVE LEARNING PERSPECTIVE
- *David H. Rakison and Gary Lupyan* (Serial No. 289, 2008).

THE GENETIC AND ENVIRONMENTAL ORIGINS OF LEARNING ABILITIES AND DISABILITIES IN THE EARLY SCHOOL YEARS
- *Yulia Kovas, Claire M. A. Haworth, Philip S. Dale, and Robert Plomin* (Serial No. 288, 2007).

THE PRESERVATION OF TWO INFANT TEMPERAMENTS INTO ADOLESCENCE
- *Jerome Kagan, Nancy Snidman, Vali Kahn, and Sara Towsley* (Serial No. 287, 2007).

CHILDREN'S QUESTIONS: A MECHANISM FOR COGNITIVE DEVELOPMENT
- *Michelle M. Chouinard* (Serial No. 286, 2007).

BEST PRACTICES IN QUANTITATIVE METHODS FOR DEVELOPMENTALISTS
- *Kathleen McCartney, Margaret R. Burchinal, and Kristen L. Bub* (Serial No. 285, 2006).

FOUNDATIONS OF SELF-AWARENESS: AN EXPLORATION THROUGH AUTISM
- *R. Peter Hobson, Gayathri Chidambi, Anthony Lee, and Jessica Meyer* (Serial No. 284, 2006).

OUR CHILDREN TOO: A HISTORY OF THE BLACK CAUCUS
OF THE SOCIETY FOR RESEARCH IN CHILD DEVELOPMENT, 1973–1997
- *Diana T. Slaughter-Defoe, Aline M. Garrett, and Algea O. Harrison-Hale* (Serial No. 283, 2006).

PARENTAL SUPPORT, PSYCHOLOGICAL CONTROL, AND BEHAVIORAL CONTROL:
ASSESSING RELEVANCE ACROSS TIME, CULTURE, AND METHOD
- *Brian K. Barber, Heidi E. Stolz, and Joseph A. Olsen* (Serial No. 282, 2005).

BEING HURT AND HURTING OTHERS: CHILDREN'S NARRATIVE ACCOUNTS
AND MORAL JUDGMENTS OF THEIR OWN INTERPERSONAL CONFLICTS
- *Cecilia Wainryb, Beverly A. Brenl, and Sonia Matwin* (Serial No. 281, 2005).

CHILDHOOD SEXUAL ASSAULT VICTIMS: LONG-TERM OUTCOMES AFTER TESTIFYING IN CRIMINAL COURT
- *Jodi A. Quas, Gail S. Goodman, Simona Ghetti, Kristen Weede Alexander, Robin Edelstein, Allison D. Redlich, Ingrid M. Cordon, and David P. H. Jones* (Serial No. 280, 2005).

THE EMERGENCE OF SOCIAL COGNITION IN THREE YOUNG CHIMPANZEES
- *Michael Tomasello and Malinda Carpenter* (Serial No. 279, 2005).

TRAJECTORIES OF PHYSICAL AGGRESSION FROM TODDLERHOOD TO MIDDLE CHILDHOOD:
PREDICTORS, CORRELATES, AND OUTCOMES
- *NICHD Early Child Care Research Network* (Serial No. 278, 2004).

CONSTRAINTS ON CONCEPTUAL DEVELOPMENT: A CASE STUDY OF THE ACQUISITION OF
FOLKBIOLOGICAL AND FOLKSOCIOLOGICAL KNOWLEDGE IN MADAGASCAR
- *Rita Astuti, Gregg E. A. Solomon, and Susan Carey* (Serial No. 277, 2004).

ORIGINS AND EARLY DEVELOPMENT OF HUMAN BODY KNOWLEDGE
- *Virginia Slaughter, Michelle Heron, Linda Jenkins, and Elizabeth Tilse* (Serial No. 276, 2004).

MOTHER–CHILD CONVERSATIONS ABOUT GENDER: UNDERSTANDING THE ACQUISITION OF ESSENTIALIST BELIEFS
- *Susan A. Gelman, Marianne G. Taylor, and Simone P. Nguyen* (Serial No. 275, 2004).

THE DEVELOPMENT OF EXECUTIVE FUNCTION IN EARLY CHILDHOOD
- *Philip David Zelazo, Ulrich Müller, Douglas Frye, and Stuart Marcovitch* (Serial No. 274, 2003).

WILEY-
BLACKWELL

MARCH/APRIL 2010 • VOLUME 81, NUMBER 2

CHILD DEVELOPMENT

Editor
JEFFREY J. LOCKMAN

Associate Editors
MARGARET BURCHINAL • JOHN COLOMBO • MICHAEL CUNNINGHAM • MARY DOZIER
ANDREW J. FULIGNI • MARY GAUVAIN • ROBERTA M. GOLINKOFF • NANCY G. GUERRA
D. GEOFFREY HALL • BETSY HOZA • MELANIE KILLEN • LOUIS J. MOSES • CYNTHIA STIFTER

CURRENT YR/VOL

Marygrove College Library
8425 West McNichols Road
Detroit, MI 48221

SRCD
Society for Research
in Child Development

Published by Wiley-Blackwell for the Society for Research in Child Development

Published by Wiley-Blackwell for the Society for Research in Child Development

Child Development

CHILD DEVELOPMENT (ISSN 0009-3920), one of three publications of the Society for Research in Child Development, Inc., is published bimonthly by Wiley Subscription Services, Inc., a Wiley Company, 111 River St., Hoboken, NJ 07030-5774. Periodical Postage Paid at Hoboken, NJ and additional offices.

Postmaster: Send all address changes to *CHILD DEVELOPMENT*, Journal Customer Services, John Wiley & Sons Inc., 350 Main Street, Malden, MA 02148-5020.

Information for Subscribers: *Child Development* is published in six issues per year. Institutional subscription prices for 2010 are: Print & Online: US$670 (US), US$938 (Rest of World), €608 (Europe), £479 (UK). Prices are exclusive of tax. Asia- Pacific GST, Canadian GST and European VAT will be applied at the appropriate rates. For more information on current tax rates, please go to www3.interscience.wiley.com/aboutus/journal_ordering_and_payment.html#Tax. The price includes online access to the current and all online back files to January 1st 1997, where available. For other pricing options, including access information and terms and conditions, please visit www.interscience.wiley.com/journal-info.

Delivery Terms and Legal Title: Prices include delivery of print journals to the recipient's address. Delivery terms are Delivered Duty Unpaid (DDU); the recipient is responsible for paying any import duty or taxes. Legal title passes to the customer on despatch by our distributors.

For more information about Wiley-Blackwell journals, including online access information, terms and conditions, and other pricing options, please visit www.interscience.wiley.com or contact our customer service department, tel: (800) 835-6770 or +1 (781) 388-8598 (Americas); +44 (0) 1865 778315 (Europe, Middle East, and Africa); or +65 6511 8000 (Asia Pacific). Subscription in 2010 includes three issues of *Monographs of the Society for Research in Child Development* and three issues of *Child Development Perspectives.* Single issues from current and recent volumes are available at the current single issue price from csjournals@wiley.com. Earlier issues may be obtained from Periodicals Service Company, 11 Main Street, Germantown, NY 12526, USA. Tel: +1 518 537 4700, Fax: +1 518 537 5899, Email: psc@periodicals.com. Volumes 37 to date are available in microfilm from ProQuest Information and Learning, Serials Customer Service Department, 300 North Zeeb Road, Ann Arbor, MI 48106; telephone: (800) 521-0600 x2873.

Electronic Access: The *Child Development* website can be found at www.srcd.org. Abstract information for this journal is electronically available at www.blackwellpublishing.com/cdev. This journal is available online at Wiley InterScience. Visit www3.interscience.wiley.com to search the articles and register for table of contents e-mail alerts.

Access to this journal is available free online within institutions in the developing world through the HINARI initiative with the WHO. For information, visit www.healthinternetwork.org

Journal Customer Services: For ordering information, claims and any enquiry concerning your journal subscription, please go to interscience.wiley.com/support or contact your nearest office:

Americas: Email: cs-journals@wiley.com; Tel: +1 781 388 8598 or 1 800 835 6770 (Toll free in the USA & Canada).
Europe, Middle East, and Africa: Email: cs-journals@wiley.com; Tel: +44 (0) 1865 778315
Asia Pacific: Email: cs-journals@wiley.com; Tel: +65 6511 8000
Japan: For Japanese speaking support, Email: cs-japan@wiley.com, Tel: +65 6511 8010 or Tel (toll-free): 005 316 50 480.
Further Japanese customer support is also available at http://www.interscience.wiley.com/support
Visit our Online Customer Self-Help available in 6 languages at http://www.interscience.wiley.com/support

Membership communication and address changes should be addressed to the SRCD Executive Office, 2950 South State Street, Suite 401, Ann Arbor, MI 48104, USA. http://www.srcd.org/

Copyright and Photocopying: Journal Compilation © 2010 The Society for Research in Child Development, Inc. All rights reserved. No part of this publication (apart from articles marked "OnlineOpen") may be reproduced, stored or transmitted in any form or by any means without the prior permission in writing from the copyright holder. Authorization to photocopy items for internal and personal use is granted by the copyright holder for libraries and other users registered with their local Reproduction Rights Organization (RRO), e.g., Copyright Clearance Center (CCC), 222 Rosewood Drive, Danvers, MA 01923, USA (www.copyright.com), provided the appropriate fee is paid directly to the RRO. This consent does not extend to other kinds of copying such as copying for general distribution, for advertising and promotional purposes, for creating new collective works or for resale. OnlineOpen articles are made available in accordance with the terms of the Creative Commons Deed, Attribution 2.5 (further details available from www.creativecommons.org), which allows Open Access dissemination of the article, but does not permit commercial exploitation or the creation of derivative works. Special requests should be addressed to PermissionsUK@wiley.com.

Production Editor: Teresa Huang (email: CDEV@wiley.com).

Advertising: Kristin McCarthy (email: kmccarthy@wiley.com).

Indexing/Abstracting: The contents of this journal are indexed or abstracted in the following: *Academic ASAP; Expanded Academic ASAP; Academic Search Elite; Academic Search Premier; Applied Social Sciences Index and Abstracts; Biomedical Reference Collection: Comprehensive Edition; CAB Abstracts; CAB Direct; Contents Pages in Education; Current Contents/Social & Behavioral Science; Current Index to Journals in Education; Cambridge Scientific Abstracts Linguistics & Language Behavior Abstracts; CatchWord; Current Index to Journals in Education; Educational Research Abstracts; Environmental Sciences & Pollution Management; e-psyche; EBSCO Sociological Collection; EBSCO Online; EBSCO Professional Development Collection; Educational Research Abstracts; Educational Resources Information Center; Health and Safety Science Abstracts; ISI Basic Social Sciences Index; InfoTrac OneFile; Ingenta; JCR Social Sciences Edition; Leisure, Recreation, and Tourism Abstracts; Online Computer Library Center FirstSearch Electronic Collections Online; Periodical Abstracts Library; Periodical Abstracts Research II; ProQuest 5000; ProQuest Social Science Plus Text; PsycINFO; Psychological Abstracts; Psychology and Behavioral Sciences Collection; Pollution Abstracts; Safety Science & Risk Abstracts; Sage Family Studies 'Abstracts; Social Services Abstracts; Social Science Citation Index; Sociological Abstracts; Student Resource Center Bronze; Student Resource Center Gold; Student Resource Center Silver; Violence & Abuse Abstracts; Wilson OmniFile V; Wilson Social Sciences, Abstracts; and Wilson Social Sciences Index*

Published by Wiley-Blackwell for the Society for Research in Child Development

Child Development

Editor

Jeffrey J. Lockman
Tulane University

Associate Editors

Margaret Burchinal, *University of North Carolina at Chapel Hill*
John Colombo, *University of Kansas*
Michael Cunningham, *Tulane University*
Mary Dozier, *University of Delaware*
Andrew J. Fuligni, *University of California, Los Angeles*
Mary Gauvain, *University of California, Riverside*
Roberta M. Golinkoff, *University of Delaware*
Nancy G. Guerra, *University of California, Riverside*
D. Geoffrey Hall, *University of British Columbia*
Betsy Hoza, *University of Vermont*
Melanie Killen, *University of Maryland*
Louis J. Moses, *University of Oregon*
Cynthia Stifter, *The Pennsylvania State University*

Managing Editor

Detra N. Davis

Editorial Assistant

Antonella Caiazza

Consulting Editors

Emma Adam
Karen E. Adolph
Nazan Aksan
William F. Arsenio
Daniel H. Ashmead
Anna Axia
Karen Bartsch
Annie Bernier
Susan Birch
Amy Booth
Heather Bouchey
Paul Boxer
Gregory Stephen Braswell
Mara Brendgen
Gene H. Brody
Keith Burt
Emily W. Bushnell
Gustavo Carlo
Stephanie M. Carlson
Ruth K. Chao
Enrique Chaux
Pablo Chavajay
Tabbye Maria Chavous
Elysia Davis
David Dickinson
Gil Diesendruck
Ruth Feldman
Lisa Gatzke-Kopp
Tamsin C. German
Rick O. Gilmore
Artin Goncu
Susan A. Graham
Elena L. Grigorenko
Laura D. Hanish
Daniel Hart
David B. Henry
Claire H. Hughes
Mutsumi Imai
Deborah J. Johnson
Roger Kobak
Joachim Krueger
Richard Lee
Charlie Lewis
Jeffrey Lidz
Tina Malti
Jelani Mandara
Gary Marcus
Ann S. Masten
Bob McMurray
Julie Meeks-Gardner
Jayanthi Mistry
Ginger A. Moore
Dianna Murray-Close
Velma McBride Murry
Letitia Naigles
Drew Nesdale
Nora S. Newcombe
David Oppenheim
Ana Papafragou
Susan M. Perez
Bradford Pillow
Eva Pomerantz
Douglas Powell
Kimberly K. Powlishta
Samuel Putnam
Suzanne M. Randolph
Robert Reeve
Kristen D. Ritchey
John J. Rieser
James L. Rodriguez
Glenn I. Roisman
Susan A. Rose
Martin D. Ruck
Jenny Saffran
Carlo Schuengel
David Schwartz
Judi G. Smetana
Alan L. Smith
Susan J. Spieker
Timothy Stickle
Dena Phillips Swanson
Adriana Umaña-Taylor
Christopher Trentacosta
Kimberly Updegraff
Marinus H. van IJzendoorn
Brian E. Vaughn
Elisa Velasquez
Lynne Vernon-Feagans
Peter Vishton
Elizabeth Votruba-Drzal
Cecilia Wainryb
Sarah Watamura
Henry M. Wellman
Mike Willoughby
Philip David Zelazo

Child Development, March/April 2010, Volume 81, Number 2, Pages 431–681

Contents

IN THIS ISSUE 431

REVIEW

The Significance of Insecure Attachment and Disorganization in the Development of Children's Externalizing Behavior: A Meta-Analytic Study
R. Pasco Fearon, Marian J. Bakermans-Kranenburg, Marinus H. van Ijzendoorn, Anne-Marie Lapsley, and Glenn I. Roisman 435

EMPIRICAL REPORTS

Not Doing What You Are Told: Early Perseverative Errors in Updating Mental Representations via Language
Patricia A. Ganea and Paul L. Harris 457

"Who Said You Could Wear My Sweater?" Adolescent Siblings' Conflicts and Associations With Relationship Quality
Nicole Campione-Barr and Judith G. Smetana 464

Categorization in 3- and 4-Month-Old Infants: An Advantage of Words Over Tones
Alissa L. Ferry, Susan J. Hespos, and Sandra R.Waxman 472

The Complex Relation Between Bullying, Victimization, Acceptance, and Rejection: Giving Special Attention to Status, Affection, and Sex Differences
René Veenstra, Siegwart Lindenberg, Anke Munniksma, and Jan Kornelis Dijkstra 480

EMPIRICAL ARTICLES

Parent Discipline Practices in an International Sample: Associations With Child Behaviors and Moderation by Perceived Normativeness
Elizabeth T. Gershoff, Andrew Grogan-Kaylor, Jennifer E. Lansford, Lei Chang, Arnaldo Zelli, Kirby Deater-Deckard, and Kenneth A. Dodge 487

Coercive and Prosocial Fathering, Antisocial Personality, and Growth in Children's Postdivorce Noncompliance
David Scott DeGarmo 503

The Tuning of Human Neonates' Preference for Speech
Athena Vouloumanos, Marc D. Hauser, Janet F. Werker, and Alia Martin 517

Early Gesture Predicts Language Delay in Children With Pre- or Perinatal Brain Lesions
Eve Sauer, Susan C. Levine, and Susan Goldin-Meadow 528

Mother–Infant Interaction During the First 3 Months: The Emergence of Culture-Specific Contingency Patterns
Joscha Kärtner, Heidi Keller, and Relindis D. Yovsi 540

"My Way or Mom's Way?" The Bilingual and Bicultural Self in Hong Kong Chinese Children and Adolescents
Qi Wang, Yi Shao, and Yexin Jessica Li 555

When Do Children Exhibit a "Yes" Bias?
Mako Okanda and Shoji Itakura 568

Lying and Truth-Telling in Children: From Concept to Action
Fen Xu, Xuehua Bao, Genyue Fu, Victoria Talwar, and Kang Lee 581

Preschoolers' Implicit and Explicit False-Belief Understanding: Relations With Complex Syntactical Mastery
Jason Low 597

Bridging Theory of Mind and the Personal Domain: Children's Reasoning About Resistance to Parental Control
Kristin Hansen Lagattuta, Larry Nucci, and Sandra Leanne Bosacki 616

Developmental Patterns in Decision-Making Autonomy Across Middle Childhood and Adolescence: European American Parents' Perspectives
Laura Wray-Lake, Ann C. Crouter, and Susan M. McHale 636

Children's Sensitivity to the Conventionality of Sources
Gil Diesendruck, Nurit Carmel, and Lori Markson 652

Children Monitor Individuals' Expertise for Word Learning
David M. Sobel and Kathleen H. Corriveau 669

OTHER

Manuscripts Accepted 680

Child Development, March/April 2010, Volume 81, Number 2, Pages 431–434

In This Issue

Quality of Care

The quality of the attachment relationship between a child and parent is important to children's development, but past research into the link between attachment and children's development has been inconsistent. **Fearon, Bakermans-Kranenburg, van IJzendoorn, Lapsley, and Roisman** (p. 435) carried out a meta-analysis, looking at 69 studies involving almost 6,000 children on the effect of the quality of child–mother attachment on children's behavior problems. Attachment insecurity is significantly associated with a higher level of children's behavior problems, particularly in boys, in clinical samples, and in studies that used observational assessments of behavior problems. Their analysis suggests that the effects of attachment are relatively persistent over time. More specifically, children who seem unable to maintain a coherent strategy for coping with separation are at greatest risk for later behavior problems and aggression.

The Helping Hand of Language

Although infants typically produce their first words sometime around their first birthdays, long before this milestone occurs they are learning a lot about language and establishing links between the language they hear and the objects and events they see. To learn how early these links emerge in infancy, **Ferry, Hespos, and Waxman** (p. 472) asked whether 3- and 4-month-olds are better able to notice categories of objects if the objects were introduced with words. Infants who heard nonsense words provided evidence of categorization, but those who heard tone sequences didn't. This tells us that even this early in development, words exert an influence on how we categorize objects that's distinct from that of nonspeech. The study provides the earliest evidence to date of a link between words and concepts; in so doing, it reinforces the guidance that parents talk to their babies, because from early on, the little ones are listening and learning.

Language allows us to learn about any topic without directly experiencing it. When new information is provided verbally about an object that isn't visible, we update our knowledge. **Ganea and Harris** (p. 457) sought to determine at what age young children do this. They found that by 19 months—almost at the onset of language production—toddlers can update based on verbal information. Even at 23 months, the updating process is error prone, but not at 30 months. This suggests that toddlers search for the object at the original location when they've been *told* about an item being moved instead of seeing it for themselves. Moreover, the children didn't make the mistakes because they couldn't encode the new information; they encoded the new information about the item's new location but failed to use it to guide their search. In showing that the ability to acquire information from words is a major cognitive advance, the study sheds light on how children use language to guide their search behavior.

Sibling Fighting and Classroom Bullying

A lot of research into fighting between siblings has focused primarily on younger children. In a study on adolescent siblings, **Campione-Barr and Smetana** (p. 464) used a new measure to examine what teens fight about and how different types of conflicts affect their relationships. The bottom line: Teen siblings argue most often about invasion of personal space (''Who said you could wear my sweater?'') and also about issues of equality and fairness (''It's my turn to sit in the front seat of the car!''). Their conflicts are frequent but not intense. Those who quarrel most often about personal space issues have poorer relationships with their siblings, with less trust and communication. The study—of more than 100 mostly White, middle-class pairs of siblings ages 13 and 15—extends our understanding of fighting between older brothers and sisters.

Veenstra, Lindenberg, Munniksma, and Dijkstra (p. 480) tackle fighting from a different perspective—bullying. They ask: Are bullies children who can't control their aggressive impulses? Not likely, according to this large-scale Dutch study. For most bullies, the pursuit of status (in this case, the desire to dominate classmates) and affection

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8102-0001

(that is, acceptance and lack of rejection) seems to foster strategic control of bullying behavior. The study, of almost 500 elementary-school children, finds that bullies generally choose to gain status by dominating their victims, but at the same time they try to reduce the chances that they'll end up on the outs with other classmates by choosing victims who are weak and not well liked by others. Gender also plays a role, with bullies being rejected by the children of the gender for whom they're a potential threat; for example, boys are rejected by boys when they bully boys and by girls when they bully girls. This study helps us understand the complex nature of bullies and victims in the elementary years.

How Divorced Dads Parent

Studies on divorced dads have found that when fathers who don't live with their children make time to be involved with them, it cuts down on problem behaviors in the youngsters. But in a new longitudinal study, **DeGarmo** (p. 503) finds that characteristics of the dads' behavior play a role in how the children do. The study took a look at 230 families with children from 4 to 11 to determine how divorced dads' parenting practices affected their children's likelihood of obeying or disobeying them, specifically taking into consideration dads' levels of social and antisocial behavior—their likeliness to have stolen, engaged in fights, be suspended from school, and so forth. For both girls and boys, when dads were antisocial, the positive effects of their monthly contact with their children were significantly reduced. The study has implications for prevention programs.

Speech Versus Nonspeech Sounds

Are infants' early listening biases for the sounds of language (as opposed to nonspeech sounds) specifically tuned to speech, or do they extend more broadly to other primate vocalizations such as rhesus monkey calls? **Vouloumanos, Hauser, Werker, and Martin** (p. 517) presented newborns and 3-month-olds with nonsense speech and the vocalizations of rhesus monkeys. Surprisingly, given that newborns have shown a preference for speech over nonspeech sounds, they found monkey calls as interesting as speech. However, the babies did prefer both monkey calls and speech to synthetic sounds. In contrast, 3-month-olds listened more to speech than to monkey vocalizations. This shows that people are born with a more broadly tuned preference that minimally includes speech and monkey vocalizations. And over the first 3 months of life, infants' listening preferences are sharpened, resulting in a species-specific preference for speech.

Brain Lesions and Language

Children with brain lesions before or around the time of birth often take longer to reach early language milestones; these delays normalize for many but persist for some. To identify which children will continue to have language delays, **Sauer, Levine, and Goldin-Meadow** (p. 528) examined whether vocabulary delays can be predicted by looking at how 18-month-old children with brain lesions used gestures such as pointing or waving good-bye when they weren't saying a lot, comparing them with typically developing children of the same age. (Gesture has been found to be a good predictor of later language abilities in typically developing children.) The children's language comprehension was tested when they were 30 months old. Findings of this longitudinal study: Gesturing at 18 months (but not early speech) predicted which children with lesions had vocabulary delays a year later, suggesting that gesture may be a tool for diagnosing persistent language delay in children with brain lesions.

Cultural Differences: Early Interaction, Self, and Discipline

Intuitively, mothers react to infants' signals quickly, normally in less than a second. This part of intuitive parenting, called contingent responsiveness, has implications for infants' later development. Earlier studies have found differences in the way mothers from different cultures react to their 3-month-olds. In a longitudinal cross-cultural study, **Kärtner, Keller, and Yovsi** (p. 540) looked at video-recorded interactions between moms and their infants prior to 3 months of age from German families (whose culture stresses independence) and from the Nso, an ethnic group in Cameroon (whose culture emphasizes interdependence). Moms in the two groups were similar in the amount of contingent responsiveness. But differences in how they reacted across their children's first 3 months suggest that moms' culture-specific beliefs and behavioral standards interact with developments in their babies' abilities to communicate (such as, for example, maintaining eye contact for longer periods). These culture-specific patterns emerged during the infants' 2nd and 3rd months of life. The results have implications for mother–child counseling programs and for those working in the field of early mother–infant interaction.

Wang, Shao, and Li (p. 555) looked at language as a representational system of culture and find that the language children speak shapes how they view themselves. They asked 125 bilingual children and adolescents (ages 8–14) from Hong Kong to recall events from their lives and describe themselves, then asked how they felt about Chinese interdependent or Western independent values. Compared with children interviewed in Chinese, children interviewed in English gave more detailed accounts of themselves and their personal experiences, focused more on unique personal attributes and qualities in describing themselves, and talked more about their own roles and perspectives in past events. Children interviewed in Chinese gave less elaborate and more socially oriented descriptions of themselves and their memories, and agreed more with Chinese interdependent values and less with Western independent values. Thus, the use of a particular language influenced whether children supported a particular cultural belief system and also shaped how they defined themselves and remembered their personal experiences. The findings offer new insights into the dynamic relations between language, culture, and the self.

Gershoff et al. (p. 487) took a cross-cultural look at discipline. Disciplining children has been studied a lot in the United States but less so in the rest of the world. Previous research has found that the meaning and effects of different forms of discipline vary by their cultural context. But this study, which sought to determine whether certain disciplinary techniques are linked with more or less child aggression regardless of the country in which the family lives, finds that disciplining children harshly is associated with negative outcomes everywhere. It also finds that for children who live in a culture in which harsh discipline is normal, the effects may not be as strong. The authors collaborated with colleagues in China, India, Italy, Kenya, the Philippines, and Thailand to look at almost three hundred 8- to 12-year-olds and their mothers. In shedding light on methods of discipline around the world, the study suggests that there's more universality in the use of particular techniques and in their associations with child behavior than previously thought.

The "Ayes" Have It

In prior research, 2- and 3-year-olds have shown a "yes" bias, that is, a tendency to say "yes" to yes–no questions pertaining to their knowledge. **Okanda and Itakura** (p. 568) looked at more than 100 Japanese 3- to 6-year-olds, finding that the reason this happens is because of both social factors (the social pressures of compliance, for example) and cognitive factors (such as that they cannot control their answer or cannot understand the conversation), and that each factor plays an important role depending on age. For example, 3-year-olds might show a yes bias automatically due to underdeveloped cognitive ability, while for older children, social reasons (such as sensitivity to interviewers' intentions) might explain their yes answers. Given that yes–no questions are the most common method used in developmental psychology, the findings suggest that researchers consider alternate methodologies for preschoolers.

Telling Lies and Falsehoods

We tend to view children as blunt truth tellers lacking in the rules of social grace, but a study of Chinese children suggests there may be something else going on. **Xu, Bao, Fu, Talwar, and Lee** (p. 581) examined whether one hundred twenty 7-, 9-, and 11-year-olds would tell lies to be polite when they received a gift they didn't like and whether children's behaviors were related to their moral understanding of lie and truth telling in a situation that called for politeness. All the children categorized untruthful statements as lies and truthful statements as the truth, regardless of whether the lies were told for prosocial or nonprosocial purposes. However, the older the children, the more they saw others' lying in polite situations less negatively and the more likely they were to tell lies in such situations themselves. Contrary to previous findings, children's moral knowledge about lying was strongly related to their actual behavior.

Low (p. 597) looked at false belief, which happens when individuals recognize that other people can have beliefs about the world that are wrong. The study looked at more than 100 preschool-age Pakeha (New Zealanders of European descent). Findings: The 3- and 4-year-olds accurately gazed in anticipation of where the main character of a story would mistakenly go to look for an object that had been moved, despite giving an incorrect verbal answer in a false-belief task. This differentiation between verbal and nonverbal responses suggests that young children have an early implicit understanding of belief before they show a mastery of certain forms of grammar that are involved in talking about the mind. In short, thought, in some core form, can exist without complex language.

Youth Autonomy

Children and adolescents identify clothing, friendship, and leisure activities as issues within the personal domain; we know that having some degree of choice over these decisions is important for both a sense of identity and mental health. **Lagattuta, Nucci, and Bosacki** (p. 616) looked at 4- to 7-year-olds' beliefs about how children would act in situations in which a parent forbids them to take part in a desired activity. Findings: From ages 4 to 7, children's predictions that characters would comply with moral rules (such as prohibitions against stealing) and feel good about doing so rose significantly, suggesting that between these ages, children showed increasing awareness of the limits to legitimate disobedience. In stark contrast, children of all ages predicted that the characters would frequently break parents' rules that intruded on the personal domain and that this disobedience would feel good, particularly when the desired activities were described as essential to the character's sense of identity. The findings suggest that children make important distinctions between different kinds of rules when reasoning about decisions and emotions.

Wray-Lake, Crouter, and McHale (p. 636) examined decision making within the family, an important way in which young people gain independence and responsibility. This longitudinal study spanning 9 years asked parents in about 200 White, European American families about their teens' decisions. They find that young people's input into decisions increased gradually from ages 9 to 14 and then surged from ages 15 to 20. Also, young people had more input into decisions about appearance, activities, schoolwork, and social life than about chores, health, and curfew. And girls, youths whom parents said were easy to supervise, and youths with better educated parents had more decision-making autonomy. Importantly, there wasn't a single, universal pattern in the development of decision making; rather, decision-making autonomy, a reflection of the development of youths' independence and responsibility, depended on what kinds of decisions youngsters faced and on their personal characteristics and family circumstances.

Trusted Sources

By age 3, children adapt their expectations about what people know about things based on their experience with others. **Diesendruck, Carmel, and Markson** (p. 652) find that children are more likely to presuppose that others share their knowledge when people have shown that they know the names of familiar objects or how they're used than when they have shown that they don't know these familiar items. Children acquire a lot of information about the world by observing and listening to others; as some of this information isn't directly taught to children but rather must be inferred by them, this study suggests that the kind of inferences preschoolers draw about the meaning of new words or functions varies depending on how knowledgeable the source of the information is. The research, of more than 100 Israeli 3- and 4-year-olds, has implications for our understanding of how children communicate with others, but also for how children learn.

Sobel and Corriveau (p. 669) considered how children learn new information, a process in which they have to rely on others as sources of information. Prior research suggests that children recognize and make inferences based on how reliable individuals are as sources of knowledge. In this study, the researchers asked whether young children register these different kinds of expertise and use that knowledge judiciously. Results: 4-year-olds gave more weight to the information provided by speakers who had information about the hidden properties of objects than to those who had more superficial knowledge. This suggests that by the time children are 4, they can pick up on cues given by speakers to decide whether the speaker is a credible source of information.

Anne Bridgman

Child Development, March/April 2010, Volume 81, Number 2, Pages 435–456

The Significance of Insecure Attachment and Disorganization in the Development of Children's Externalizing Behavior: A Meta-Analytic Study

R. Pasco Fearon
University of Reading

Marian J. Bakermans-Kranenburg
and Marinus H. van IJzendoorn
University of Leiden

Anne-Marie Lapsley
Barnet, Enfield & Haringey Mental Health NHS Trust

Glenn I. Roisman
University of Illinois at Urbana-Champaign

This study addresses the extent to which insecure and disorganized attachments increase risk for externalizing problems using meta-analysis. From 69 samples (N = 5,947), the association between insecurity and externalizing problems was significant, d = 0.31 (95% CI: 0.23, 0.40). Larger effects were found for boys (d = 0.35), clinical samples (d = 0.49), and from observation-based outcome assessments (d = 0.58). Larger effects were found for attachment assessments other than the Strange Situation. Overall, disorganized children appeared at elevated risk (d = 0.34, 95% CI: 0.18, 0.50), with weaker effects for avoidance (d = 0.12, 95% CI: 0.03, 0.21) and resistance (d = 0.11, 95% CI: −0.04, 0.26). The results are discussed in terms of the potential significance of attachment for mental health.

Although the significance of the parent–child relationship was recognized by scientists and clinicians since the earliest days of formal psychological inquiry (e.g., Baldwin, 1895; Freud, 1908; James, 1890), two major advances occurred in the 1960s and 1970s that created lasting legacies for the study of human development. John Bowlby's (1969) theory of parent–child attachment was revolutionary in the way it integrated evolutionary, biological, developmental, and cognitive concepts into a unified account of human attachment behavior. This remarkable achievement paved the way for the scientific study of attachment, largely because it created a conceptual framework for developing testable hypotheses about causal influences, developmental processes, and expected long-term consequences of attachment for mental health (Bretherton, 1997). Critical among these novel contributions were the clear characterization of the proximal behavioral functions associated with attachment and their interplay with other biologically significant behavioral systems, the use of comparative evidence as crucial sources of theory-development, and the concept of an internal working model as a framework for understanding continuities in attachment behavior across context and over time. The notion that the quality or organization of attachment behavior in early infancy or childhood might have implications for later socioemotional development and mental health is arguably one of attachment theory's most well-known and contested predictions (Lamb, Thompson, Gardner, Charnov, & Estes, 1984; Rutter, 1995). Nowhere is this issue more significant than in the domain of aggression and externalizing behavior problems, where the social costs are substantial (Loeber & Hay, 1997).

Researchers working in the attachment field, following Bowlby and others, have considered a number of mechanisms that might explain why attachment experiences in early life might be associated with later adaptation and mental health. Several theorists have suggested that the role of attachment may center on the way in which

---

Support from the Netherlands Organization for Scientific Research to Marinus van IJzendoorn (NWO SPINOZA award) and to Marian Bakermans-Kranenburg (NWO VIDI grant) is gratefully acknowledged. The study was also supported in part by a grant from the National Science Foundation to Glenn I. Roisman (BCS-0720538).

Correspondence concerning this article should be addressed to R. Pasco Fearon, School of Psychology and Clinical Language Sciences, University of Reading, 3 Earley Gate, Whiteknights, Reading RD6 6AL, United Kingdom. Electronic mail may be sent to p.fearon@reading.ac.uk.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8102-0002

children respond to sources of threat and challenge, and the extent to which children are able to draw on parental support and comfort as a means of coping (Kobak, Cassidy, Lyons Ruth, & Ziv, 2005). Secure children, it is maintained, have had repeated experiences of a caregiver who is responsive when support and proximity are needed and expect the caregiver(s) to be available and comforting when called upon. In contrast, children with insecure attachment relationships may have had experiences in which bids for proximity have been discouraged, rejected, or inconsistently responded to and rely more heavily on secondary coping processes to deal with stress and challenge.

Developmental continuities between the organization of the attachment relationship and functioning beyond it (in time or space) have generally been conceptualized with reference to the internal working models construct. This important conceptual heuristic is thought of as a set of organized cognitive–affective psychological structures that organize thinking, feeling, and behavior vis-à-vis the attachment figure as a potential haven of safety and comfort in times of stress (Bretherton, 1995; Main, Kaplan, & Cassidy, 1985). These models are thought to become generalized over time and influence functioning in wider interpersonal relationships across the life span and form the basis of a generalized sense of the self as worthy of love and care and others as available and responsive (Cassidy, 1988; Sroufe, Egeland, Carlson, & Collins, 2005a). In addition to this primary explanatory construct, several other factors (possibly related to internal working models) have been discussed as potential mediators between a secure attachment relationship and lowered risk for mental health problems generally and externalizing problems more specifically, such as (a) a developing sense of self-confidence through repeated experiences of support and comfort and through effective exploration of the environment (Goldberg, 1997), (b) generalized positive social expectations (as opposed to mistrust and perceived hostility; see Dodge & Coie, 1987), (c) the socialization of moral emotions and values within a secure attachment relationship (Kochanska, 1997; van IJzendoorn, 1997), (d) modeling of prosocial behavior by a sensitive caregiver (Guttmann-Steinmetz & Crowell, 2006), (e) continuity in the quality and supportiveness of ongoing parental care (Lamb et al., 1984), and (f) the capacity for effective emotion regulation (e.g., Cassidy, 1994). Other possibilities exist that have been given less attention, such as the social modulation of biological systems mediating stress and arousal regulation (e.g., Suomi, 2003; Weaver et al., 2004).

Although Bowlby's work was highly influential, the development of a standardized procedure for the systematic study of attachment behavior, as observed in a naturalistic setting, was a major further step forward in the establishment of an empirical knowledge base concerning the developmental significance of attachment (Ainsworth, Blehar, Waters, & Wall, 1978; Sroufe, 1983). Ainsworth et al.'s (1978) Strange Situation procedure (SSP) has become one of the most widely used—if not the most widely used—standardized lab assessment of early childhood behavior based on direct observation and represents a paradigm example of how to systematically study naturally occurring behavior in quasi-naturalistic contexts. The identification of individual differences in patterns of reunion behavior following separation in the SSP triggered a program of research studies aimed at uncovering their developmental antecedents and sequelae (Belsky & Isabella, 1988; Schneider-Rosen & Rothbaum, 1993). Despite the accumulation of an impressive volume of data over the years, the picture that unfolded regarding the developmental consequences of attachment has proved complex and often contradictory, particularly in the domain of mental health and psychopathology (Goldberg, 1997).

One of the earliest and most influential longitudinal studies of the psychosocial outcomes of children observed in the SSP in infancy was launched in Minnesota by Byron Egeland, Alan Sroufe, and colleagues (see Sroufe, Egeland, Carlson, & Collins, 2005a, 2005b). Erickson, Sroufe, and Egeland (1985) followed their relatively large sample of high-risk infants from 12 months to preschool and collected extensive assessments of children's behavior using observer and teacher ratings of social competence, ego control, peer confidence, and externalizing behavior problems in the school setting. Most importantly for the current purposes, secure children scored lower than insecure children on assessments of behavior problems and avoidant children stood out as being particularly at risk, a finding echoed in several later studies (Burgess, Marshall, Rubin, & Fox, 2003; Goldberg, Gotowiec, & Simmons, 1995; Munson, McMahon, & Spieker, 2001). Interestingly, a later report from the Minnesota study at Grades 1–3 (Renken, Egeland, Marvinney, Mangelsdorf, et al., 1989) found associations between attachment insecurity and externalizing problems in boys but not girls (see also Lewis, Feiring, McGuffog, & Jaskir, 1984).

In a manner that was to become somewhat characteristic of the topic, the first reports of these findings were followed immediately by a nonreplication. In the same monograph, Bates, Maslin, and Frankel (1985) reported on a longitudinal follow-up of 120 infants who had previously been observed in the SSP at 12 months and found no association between attachment security and parent reports of externalizing behavior problems at age 3. Any number of methodological factors could be considered when interpreting these early, apparently contradictory findings. Notably, Erickson et al.'s (1985) study excluded cases that were not stable in terms of attachment classifications between 12 and 18 months, while Bates et al. (1985) only collected attachment data at 12 months. There are obvious reasons why the stability of attachment might be a factor in its predictive power. Furthermore, Sroufe et al.'s study was drawn from a substantially more impoverished population than the Bates et al. study, which was predominantly middle class. A number of authors have argued that attachment security should be thought of as an interactive risk factor that is more significant when other psychosocial stressors are present in the family ecology (Belsky & Fearon, 2002; Kobak et al., 2005). The two studies also employed different outcome measures (teacher vs. parent report), which in turn may index contextual differences in the expression of externalizing behavior or in the validity of the assessments.

Thus, even in the earliest phase of research into the longitudinal outcomes of attachment security and insecurity, positive findings, negative findings, and interactions emerged in almost equal measure. A similar mix of results emerged from later studies conducted in the 1980s and early 1990s. With such a complex pattern of study outcomes, narrative reviews took diverging positions regarding the status of the evidence for an association between attachment and children's behavior problems (see Belsky & Nezworski, 1988).

The identification of disorganized attachment (Main & Solomon, 1986, 1990) led to renewed interest in the potential for attachment to predict robustly externalizing behavior problems (Carlson, 1998; Lyons Ruth, Alpern, & Repacholi, 1993; Moss, Cyr, & Dubois Comtois, 2004). These seemingly inexplicable, contradictory, and fragmentary behaviors observed during the SSP are considered by many to represent relational processes at special risk for psychopathology, particularly in the domain of childhood aggression (Liotti, 1992; Lyons Ruth, Zeanah, & Benoit, 2003; Main & Morgan, 1996; Moss, Cyr, et al., 2004). Several authors have outlined hypotheses regarding the mechanisms by which disorganized attachment may lead to aggression, with considerable attention focusing on states of emotional dysregulation and dissociative processes that may block the person's awareness of his or her violent actions (e.g., Fonagy, 2004; Liotti, 1992; Solomon & George, 1999). Subsequently, a sizable body of evidence emerged that was consistent with the view that disorganized attachment may be associated with increased risk for externalizing behavior problems and aggression. A meta-analysis of 12 studies carried out in 1999 (van IJzendoorn, Schuengel, & Bakermans-Kranenburg, 1999) showed that this association was robust, with a mean effect size of $r = .29$ ($N = 734$). Nevertheless, since 1999 a significant number of new studies have been conducted, including the largest ever longitudinal study of attachment (with more than 1,000 participants), the National Institute of Child Health and Human Development (NICHD) Study of Early Child Care and Youth Development, which failed to find strong evidence of greater externalizing behavior problems in disorganized children (Belsky & Fearon, 2002; NICHD Early Child Care Research Network, 2006).

The question of whether attachment insecurity plays a causal role in the development of externalizing psychopathology is a vital one for the field, but there is clearly little case for causality if there is no association. With the sheer volume, range, and diversity of studies that have examined the association between attachment security and children's externalizing behavior problems, it has become virtually impossible to provide a clear narrative account of the status of the evidence concerning this critical issue in developmental science. Given that sample variability around an effect of zero can lead to false positives, and sampling variability around a positive effect can lead to false negatives, the question of whether the existing evidence is consistent with positive association is critical for a full appreciation of the predictive significance of attachment for later externalizing behavior problems. Meta-analysis provides a structured, principled methodology for resolving—within limits—these essential scientific questions. In the current study, we analyzed over 60 independent studies that have conducted assessments of attachment security and insecurity using standardized observational tools and related them to measures of children's externalizing behavior problems. In line with expectations derived from the literature, we set out to test several hypotheses, namely, that (a) attachment insecurity, in particular avoidant attachment, would be significantly associated with externalizing behavior problems; (b) stronger effects would be found in low-socioeconomic-status (SES) samples than high-SES samples; (c) stronger associations would be found in boys than in girls; and (d) attachment disorganization would predict externalizing problems more strongly than avoidance or insecurity generally. We added to these a focus on whether effects of attachment-related variation were moderated by age of assessment of externalizing problems. The claim that early attachment has enduring—rather than merely transient—effects on development requires that the magnitude of such associations are not reduced to nil over time. Finally, we also examined a range of relevant methodological factors that might account for systematic between-study variability in effect sizes, including the method of assessment of attachment and the type and context of outcome measurement.

Method

Literature Search

We systematically searched the electronic databases PsychInfo, Web of Science, MEDLINE, Science Citation Index Expanded, Social Sciences Citation Index, and Art & Humanities Citation Index with the key words externalizing, aggressi*, conduct, psychopathology, opposition*, competence, social functioning, prosocial, antisocial, anti-social, behavior problem*, behaviour problem* in the title or abstract (the asterisk indicates that the search contained the word or word fragment). This large set was narrowed down by adding the constraint that the papers must also contain the word attachment and child* or infan* in the title or abstract. This search returned over 1,200 articles in each of the databases. Two further separate search-narrowing strategies were adopted, which yielded two partially overlapping study sets. In one search, we targeted empirical studies by requiring the abstract to contain the words sample or *N*. In another, we required that at least one of the words secur*, avoidan*, resistan*, or disorgani* appeared in the title or abstract. In each case, this reduced the former search by around 60%. When these two sets of search results were merged, this resulted in 856 candidate articles. These were subjected to abstract review in the first instance, from which a large number of clearly irrelevant articles were discarded (e.g., nonempirical papers, studies not involving children). A further 115 articles remained. These were examined individually by the authors according to criteria described next. Second, the reference lists of the collected empirical papers and influential reviews were searched for relevant studies (e.g., Kobak et al., 2005). Third, data sets available to the authors since they were in the public field (NICHD Study of Early Child Care and Youth Development) or part of their ongoing research (SCRIPT; Van Zeijl et al., 2006) were analyzed with regard to the associations between attachment and externalizing behavior.

Studies were included if they reported on the relation between attachment and externalizing or aggressive behavior in children 12 years of age or younger. Externalizing behavior was defined as aggression, oppositional problems, conduct problems or hostility (either alone or in combination), as indicated in the descriptions provided in the method sections of the respective articles. Studies that did not differentiate between externalizing and internalizing problems (e.g., just total problems score of the CBCL) were excluded. Externalizing or aggressive behavior was assessed using observation (e.g., Matas, Arend, & Sroufe, 1978; Turner, 1991), questionnaires (CBCL, PBQ) or clinical interviews (e.g., Speltz, DeKlyen, & Greenberg, 1999), completed by parents (e.g., Aviezer, Sagi, Resnick, & Gini, 2002), teachers (e.g., Egeland & Heister, 1995), or clinicians or trained observers (e.g., Turner, 1991). We restricted the review to studies using observational measures of attachment, such as the SSP (Ainsworth et al., 1978), the Cassidy & Marvin Preschool Attachment system (Cassidy, Marvin, and The MacArthur Working Group on Attachment, 1989), the Attachment Q-Sort (AQS; Waters & Deane, 1985), and the Main and Cassidy system (Main & Cassidy, 1988). In cases where more than one attachment assessment was employed (e.g., the SSP followed by Cassidy & Marvin at a later age), the earliest attachment assessment was selected. We did not include studies that reported on representational measures of attachment (e.g., Attachment Story Completion Task; Verschueren & Marcoen, 1999). When intervention studies were identified, we only included data from the non-treated control sample (e.g., Lieberman, Weston, & Pawl, 1991). Only one study that met our entry criteria also reported on outcome data for father–child attachment security (Aviezer et al., 2002). As this would not allow a meaningful comparison of effect sizes between mother and father attachment, this study was excluded. The meta-analyses reported herein therefore only pertain to mother–child attachment. Also noteworthy is the fact that only four studies were based on samples from predominantly minority-ethnic communities.

Several studies presented data on (partly) overlapping samples, such as Shaw and colleagues (Shaw, Owens, Vondra, Keenan, & Winslow, 1996; Shaw & Vondra, 1995) and the studies reported by Moss and colleagues (Moss, Bureau, Cyr, Mongeau, & St Laurent, 2004; Moss, Cyr, et al., 2004; Moss, Parent, Gosselin, & Rousseau, 1996). Because participants cannot be included in a meta-analysis more than once, the papers that reported on the largest groups of participants were included in our meta-analysis (e.g., Moss, Bureau, et al., 2004). In total, after excluding reports involving overlapping samples, we found 53 studies that yielded 69 independent samples that could be included in our meta-analyses, with sample sizes ranging from 26 to 1,075 (see Table 1). In many cases, outcome statistics were only presented for the avoidant and resistant classifications combined, or indeed for the resistant, avoidant, and disorganized cases combined. Consequently, we focused our primary analyses on the overall contrast between security and insecurity, with insecurity represented by the avoidant, resistant and (in the cases where disorganization had been coded) disorganized classifications. In these analyses, we also tested whether it made a difference to the overall effect size for security if disorganization had been coded. In addition, a number of studies used the AQS to measure attachment security, which does not yield data on the different subtypes of insecurity. As a result, these studies only appear in the meta-analyses involving the overall contrast between security and insecurity. Subsequently, we also extracted more focused contrasts targeting specific insecure categories from the smaller set of studies where these could be identified. The numbers of studies involved in these subanalyses are indicated in the text.

Coding System

We used a structured coding system for assessing the characteristics of the samples and their study designs. The measurement of attachment was coded straightforwardly, as all studies included one of several well-known attachment assessments (SSP, AQS, Preschool Attachment Assessment, Cassidy & Marvin; Main & Cassidy). In each case, the coder extracted effect sizes at the level of the individual attachment classification where possible (i.e., A, B, C, and D). In addition to a number of background variables like year of publication and data source (journal, book chapter, unpublished data), we coded several important potential moderators related to the sample: gender (% male), SES (high/middle vs. low), and clinical status (clinical-child, clinical-parent, nonclinical). Where the gender composition of the sample was not precisely reported we assumed a 50% split. Furthermore, when SES was not noted, a default of high/middle class was recorded (this occurred in five cases). Clinical status was recorded if either the parent or the target child was identified as having a clinical diagnosis or if they had been selected using a clinical cutoff score on a validated instrument. In addition to the measure used to assess attachment, four other design characteristics were coded: (a) age of attachment assessment, (b) age of externalizing assessment, (c) type of outcome measure, and (d) observer or reporter of externalizing behavior. When measurements of externalizing problems at different points in time were reported, we selected the first measurement. When measurements at different points in time were in some way merged (averaged or a trajectory extracted), we recorded the midpoint of the range as the age of the externalizing assessment (e.g., Keller, Spieker, & Gilchrist, 2005). To assess intercoder reliability, 20 randomly selected studies were coded by two coders. The agreement between the coders across the moderator variables was 97% (correlations between continuous moderators were > .95).

Meta-Analytic Procedures

A number of studies reported results separately for boys and girls and four studies reported on samples involving only boys or only girls. In these cases, we calculated separate effect sizes for each gender, and the subsamples were treated as independent outcomes in the analyses. When multiple measures of aggression or externalizing behaviors were used within one study (e.g., Solomon, George, & De Jong, 1995), we selected the outcome for externalizing behavior for the primary set of studies. In a separate set of meta-analyses, we tested whether outcomes were different when aggression outcomes were selected (see the following).

Statistical Analyses

Four sets of meta-analyses were conducted, one for the relation between attachment insecurity and externalizing or aggressive behavior, one for the relation between avoidance and externalizing or aggressive behavior, one for the relation between resistance and externalizing or aggressive behavior, and one for the relation between attachment disorganization and externalizing or aggressive

Table 1
Sample Characteristics for Primary Set of Studies

Source	Subsample description	Attachment measure	Outcome	*N*
Anan & Barnett (1999)		CassMarv	CBCL	56
Aviezer et al. (2002)		SSP	CBCL	63
Bakermans-Kranenburg & van IJzendoorn (2006)	Boys subsample	SSP	CBCL	23
	Girls subsample	SSP	CBCL	24
Bates & Bayles (1988)	Boys subsample	SSP	CBCL	28
	Girls subsample	SSP	CBCL	27
Booth et al. (1991)	High-risk subsample	SSP	Obs	20
	Low-risk subsample	SSP	Obs	16
Booth et al. (1994)		SSP	CBCL	69
Burgess et al. (2003)		SSP	CBCL	140
Carlson (1998)	Minnesota study: Effect size for D versus non-D only	SSP	CBCL	78
Cicchetti et al. (1998)		AQS	CBCL	128
Cohn (1990)	Boys subsample	SSP	Other	34
	Girls subsample	SSP	Other	46
DeMulder et al. (2000)	Boys subsample	AQS	Other	51
	Girls subsample	AQS	Other	43
DeVito & Hopkins (2001)		PAA	CBCL	58
Edwards et al. (2006)	Clinical fathers subsample	SSP	CBCL	82
	Nonclinical fathers subsample	SSP	CBCL	94
Egeland & Heister (1995)	Minnesota study: B versus non-B effect size	SSP	CBCL	64
Fagot & Leve (1998)		SSP	CBCL	136
R. P. Fearon, unpublished data	Twin sample	SSP	CBCL	27
Gilliom et al. (2002)		SSP	CBCL	189
Goldberg et al. (1995)	Cystic fibrosis sample	SSP	CBCL	40
	Coronary heart disease sample	SSP	CBCL	54
	Nonclinical sample	SSP	CBCL	51
Goldberg et al. (1990)		SSP	Other	69
Greenberg et al. (1991)	ODD + Controls	CassMarv	Diag	50
Howes et al. (1994)		SSP	Other	74
Hubbs Tait et al. (1994)		SSP	CBCL	44
Keller et al. (2005)		SSP	CBCL	169
Klein Velderman et al. (2006)		SSP	CBCL	26
Lewis et al. (1984)	Boys subsample modified SSP	SSP	CBCL	51
	Girls subsample modified SSP	SSP	CBCL	57
Lieberman et al. (1991)	Intervention— control sample only	SSP	Obs	52
Lyons Ruth et al. (1993)		SSP	Other	62
Madigan et al. (2007);.		SSP	CBCL	64
Marchand & Hock (1998)		AQS	CBCL	46
Matas et al. (1978)		SSP	Obs	48
Moss, Cyr, et al. (2004)		CassMarv	Other	220
Munson et al. (2001)		SSP	CBCL	101
NICHD	Boys high-SES subsample	SSP	CBCL	490
	Boys low-SES subsample	SSP	CBCL	55
	Girls high-SES subsample	SSP	CBCL	460
	Girls low-SES subsample	SSP	CBCL	70
F. D. Pannebakker, unpublished data		SSP	CBCL	115
Perez Corres (2006)		SSP	CBCL	51
Pierrehumbert et al. (2000)		SSP	CBCL	40
Radke Yarrow et al. (1995)		SSP	Diag	95

Table 1
Continued

Source	Subsample description	Attachment measure	Outcome	*N*
Rothbaum et al. (1995)		SSP	CBCL	32
Schmidt et al. (2002)		AQS	CBCL	49
C Schuengel, M. J. Bakermans-Kranenburg, & M. H. van IJzendoorn, unpublished data		SSP	CBCL	38
Seifer et al. (2004)		SSP	CBCL	732
Shaw et al. (1996)		SSP	CBCL	77
Smeekens et al. (2007)		SSP	CBCL	105
Solomon et al. (1995)		MC	Other	40
Speltz et al. (1990)		CassMarv	Diag	50
Speltz et al. (1999)		CassMarv	Diag	160
Stams et al. (2002)		SSP	CBCL	155
Suess et al. (1992)		SSP	Obs	35
Turner (1991)	Boys subsample	CassMarv	Obs	18
	Girls subsample	CassMarv	Obs	22
van IJzendoorn et al. (1992)		SSP	Other	68
Van Zeijl et al. (2006)	1-year SSP boys	CassMarv	CBCL	25
	1-year SSP girls	CassMarv	CBCL	18
	2-year SSP boys	CassMarv	CBCL	28
	2-year SSP girls	CassMarv	CBCL	10
	3-year SSP boys	CassMarv	CBCL	18
	3-year SSP girls	CassMarv	CBCL	16
Vondra et al. (2001)		SSP	CBCL	165
Weiss & Seed (2002)		AQS	CBCL	110
Wood et al. (2004)		AQS	CBCL	37

Note. AQS = Waters and Deane (1985) Attachment Q-Set; CassMarv = Cassidy and Marvin/MacArthur Preschool Attachment Coding System; SES = socioeconomic status; SSP = Strange Situation procedure; PAA = Crittenden Preschool Attachment Assessment; MC = Main and Cassidy Age 6 Scoring System; Obs = externalizing directly observed; CBCL = Child Behavior Checklist; Other = other externalizing questionnaire; Diag = clinical diagnosis.

behavior. The meta-analyses were performed using the Comprehensive Meta-Analysis (CMA) program (Version 2; Borenstein, Rothstein, & Cohen, 2005). For each study, an effect size (*d*) was calculated as the standardized difference between the two pertinent groups (e.g., secure vs. insecure). In those cases where continuous attachment scores were correlated with externalizing scores (e.g., when the study reported on the AQS), we recomputed the statistic into Cohen's *d* (see Mullen, 1989; Mullen & Rosenthal, 1985, chap. 6, for the formulae for transformation of various statistics into Cohen's *d*). Effect sizes indicating a positive relation between externalizing behavior and insecurity, avoidance, and disorganization, respectively (higher levels of externalizing behavior in the insecure, avoidant, or disorganized group compared to the reference group), were given a positive sign. Thus, a positive combined effect for the set of studies comparing disorganized children with secure children on externalizing behaviors would mean that across these studies the level of externalizing behaviors in disorganized children was higher than in secure children. In the main analyses, we compared externalizing behaviors of the children in each attachment classification with all other classifications combined. In an additional analysis on a smaller set of studies with pertinent data, we also compared each classification with the secure classification as the most "pure" reference category.

Using CMA, combined effect sizes were computed. Significance tests and moderator analyses were performed through fixed or random effects models, depending on the homogeneity of the study outcomes. Fixed effects models are based on the assumption that effect sizes observed in a study estimate the corresponding population effect with random error that stems only from the chance factors associated with subject-level sampling error in that study (Lipsey & Wilson, 2001; Rosenthal, 1991). This assumption is not made in random effects models (Hedges & Olkin, 1985). Random effects models allow for the possibility that there are random differences between studies that are associated with variations in procedures, measures, and settings that go beyond subject-level sampling error and thus point to different study populations (Lipsey & Wilson, 2001). To test the homogeneity of the overall and specific sets of effect sizes, we computed Q statistics (Borenstein et al., 2005). In addition, we computed 95% confidence intervals (CIs) around the point estimate of each set of effect sizes. When the set was homogeneous, CIs were based on fixed estimates. In cases where there was heterogeneity across studies, we based CIs on random estimates. Q statistics and p values were also computed to assess differences between combined effect sizes for specific subsets of studies grouped by moderators. Again, fixed effects model tests were used in the case of homogeneous sets of outcomes, and more conservative random effects model tests were used in the case of heterogeneous outcomes. In the present study, random models were tested unless otherwise specified. Contrasts were only tested when at least two of the subsets consisted of at least four studies (Bakermans-Kranenburg, van IJzendoorn, & Juffer, 2003).

When the children in two sets of studies (partially) overlapped (e.g., some studies reported both aggression and externalizing, and we wanted to compare the combined effects for aggression and externalizing), it was impossible to compare directly effect sizes across these sets. We computed 85% CIs for the point estimates of the combined effect sizes in the two sets: Nonoverlapping 85% CIs indicate a significant difference between combined effect sizes (Bakermans-Kranenburg et al., 2003). This approach of comparing 85% CIs served as a conservative significance test (Goldstein & Healy, 1995). We used the "trim-and-fill" method (Duval & Tweedie, 2000a, 2000b) to calculate the effect of potential data censoring (or publication bias) on the outcome of the meta-analyses. Using this method, a funnel plot is constructed of each study's effect size against the sample size or the standard error (usually plotted as $1/SE$, or precision). It is expected that this plot has the shape of a funnel because studies with smaller sample sizes (larger standard errors) have increasingly larger variation in estimates of their effect size as random variation becomes increasingly influential, whereas studies with larger sample sizes have smaller variation in effect sizes (Duval & Tweedie, 2000b; Sutton, Duval, Tweedie, Abrams, & Jones, 2000). The plots should only be shaped like a funnel if no data censoring is present. However, as smaller or nonsignificant studies are less likely to be published (the "file-drawer" problem; Mullen, 1989), studies in the bottom left-hand corner of the plot are often omitted (Sutton et al., 2000). In our meta-analyses, the k right-most studies considered to be symmetrically unmatched were trimmed. The trimmed studies can then be replaced and their missing counterparts imputed or "filled" as mirror images of the trimmed outcomes. This then allows for the computation of an adjusted overall effect size and CI (Gilbody, Song, Eastwood, & Sutton, 2000; Sutton et al., 2000).

For each meta-analysis, we also calculated the number of studies with average sample size and nonsignificant outcome that would be required to bring the combined effect size of the meta-analysis to a nonsignificant level (fail-safe number; Mullen, 1989). Rosenthal (1991, p. 106) suggested that a fail-safe number of more than $5k + 10$ (k = number of studies included) may be considered a general criterion for robustness.

For each study, Fisher's Z scores were computed as well-distributed equivalents for the effect size d, and the Z scores were standardized to test for outliers. No outliers (standardized Z values smaller than -3.29 or larger than 3.29; Tabachnik & Fidell, 2001) were found for study effect sizes.

Results

Is Insecure Attachment Associated With More Externalizing Problems?

The first set of meta-analyses concerned the difference in externalizing behaviors between children rated as secure versus children classified as insecure. In 69 studies including $N = 5{,}947$ participants, the association between attachment security and externalizing behaviors was reported. Any study assessing attachment and externalizing was included in this total set, regardless of type of measures used. If disorganized attachment was assessed, this category was included in the insecure group. In this overall set, we found a significant combined effect size of $d = 0.31$ (see Table 2). Children rated as insecure showed higher levels of externalizing behaviors than children rated as secure. The effect size was modest but robust, as more than 1,700 studies with null results (fail-safe number) would be needed to reduce this effect to nonsignificance. Further support for the absence

Table 2

Insecure Attachment and Externalizing Behavior: Total Set and Set of SSP Studies

	k	N	*d*	Confidence interval 95%	Homogeneity Q	Contrast Q^a	Contrast *p*
Total set	69	5,947	0.31**	0.23, 0.40	135.18**		
Clinical						5.48	.02
Nonclinical	56	4,812	0.26**	0.17, 0.35	89.06**		
Clinical	13	1,135	0.49**	0.32, 0.66	29.98**		
Gender						8.49	.01
Boys	14	1,210	0.35**	0.17, 0.54	32.41**		
Girls	12	907	−0.03	−0.16, 0.11	13.93		
Mixed	43	3,830	0.36**	0.26, 0.46	69.79**		
SES						0.05	.82
Low	14	1,801	0.25**	0.15, 0.35	15.49		
Middle/high	55	4,146	0.31**	0.21, 0.41	119.69**		
Measure						14.49	< .01
SSP	43	4,488	0.18**	0.12, 0.24	57.31		
AQS	7	464	0.70**	0.51, 0.90	5.57		
CassMarvin	12	708	0.37**	0.16, 0.57	32.18**		
Other	7	287	0.39**	0.14, 0.64	10.93		
Observer ext						10.27^a	.02
Mother	44	4,129	0.22**	0.12, 0.32	72.37**		
Teacher	10	922	0.30**	0.17, 0.44	15.72		
Observed	7	211	0.58**	0.30, 0.86	4.12		
Other	6	425	0.62**	0.35, 0.89	15.83**		
Combined	2	260	0.45**	0.20, 0.71	2.21		
SSP studies							
All SSP studies	43	4,488	0.18**	0.12, 0.24	57.31		
Clinical						0.03	.86
Nonclinical	36	3,899	0.22**	0.13, 0.31	55.33*		
Clinical	7	589	0.21*	0.04, 0.38	1.83		
Gender						13.08	< .01
Boys	6	836	0.18*	0.04, 0.32	4.27		
Girls	6	753	−0.06	−0.21, 0.09	8.52		
Mixed	31	2,899	0.24**	0.17, 0.32	31.44		
SES						0.90	.34
Low	12	1,635	0.22**	0.12, 0.32	11.47		
Middle/high	31	2,853	0.19**	0.09, 0.29	44.87*		
Observer ext						8.91^a	.01
Mother	31	3,499	0.13**	0.06, 0.20	39.35		
Teacher	4	463	0.22*	0.03, 0.40	0.79		
Observed	5	171	0.61**	0.29, 0.92	0.80		
Other	1	95	0.32	−0.10, 0.73			
Combined	2	260	0.45**	0.20, 0.71	2.21		

Note. SES = socioeconomic status; SSP = Strange Situation procedure; AQS = Attachment Q-Sort; CassMarvin = Cassidy and Marvin/MacArthur Preschool Attachment Coding System.
^{a}Subgroup with $k < 4$ excluded from contrast.
$*p < .05$. $**p < .01$.

of the file-drawer problem was evident through the trim-and-fill approach, showing that only five studies had to be trimmed and filled, with a resulting significant combined effect size of $d = 0.27$ (95% CI: 0.18, 0.36).

As the total set of studies for secure versus insecure attachment was heterogeneous (see Table 3), we looked for significant moderators that might account for between-study variability in outcome. Gender appeared to be a significant moderator (see Table 2). In samples with only girls, the combined effect size was $d = -0.03$ (*ns*), whereas in the samples with only boys and in the mixed samples (i.e., with boys and girls) the combined effect sizes were

Table 3

Avoidant Attachment and Externalizing Behavior: Total Set and Set of SSP Studies

	k	*N*	*d*	Confidence interval 95%	Homogeneity *Q*	Contrast *Q*[a]	Contrast *p*
Total set	34	3,675	0.12**	0.03, 0.21	40.54		
Clinical						0.93	.33
Nonclinical	28	3,273	0.10*	0.01, 0.20	35.62		
Clinical	6	402	0.22*	0.01, 0.43	3.98		
Gender						1.24	.54
Boys	9	890	0.19*	0.03, 0.36	5.32		
Girls	7	763	0.13	−0.09, 0.34	8.45		
Mixed	18	2,022	0.08	−0.04, 0.20	25.53		
SES						2.42	.12
Low	5	1,001	−0.01	−0.20, 0.17	6.81		
Middle/high	29	2,674	0.15**	0.06, 0.25	31.31		
Measure						0.00[a]	.98
SSP	25	3,054	0.13**	0.03, 0.23	26.01		
CassMarvin	7	523	0.13	−0.11, 0.37	12.78*		
Other	2	98	−0.20	−0.69, 0.28	0.01		
SSP studies							
All SSP studies	25	3,054	0.13**	0.03, 0.23	26.01		
Gender						0.03	.98
Boys	5	647	0.12	−0.08, 0.33	1.09		
Girls	6	753	0.16	−0.06, 0.38	1.77		
Mixed	14	1,654	0.15*	0.01, 0.33	23.03*		
SES						2.95	.10
Low	5	1,001	−0.01	−0.20, 0.17	6.81		
Middle/high	20	2,053	0.18**	0.07, 0.29	16.25		

Note. SES = socioeconomic status; SSP = Strange Situation procedure; CassMarvin = Cassidy and Marvin/MacArthur Preschool Attachment Coding System.

[a] Subgroups with $k < 4$ excluded from contrast.

$^{*}p < .05$. $^{**}p < .01$.

significant, $d = 0.35$ and $d = 0.36$, respectively, and significantly different from the null effect in the girls-only samples. Within the mixed samples, the percentage of males in the study tended to be positively related to the magnitude of the attachment-externalizing problems effect size (slope = .013, $p = .11$). Clinical samples also showed a significantly larger combined effect size ($d = 0.49$) than nonclinical samples ($d = 0.26$). This moderator effect was not dependent on gender differences as only two of the clinical samples consisted of boys only (Greenberg, Speltz, Deklyen, & Endriga, 1991; Speltz et al., 1999). Contrary to expectations, SES was not a significant moderator (see Table 2).

The type of assessment of externalizing behaviors and of attachment appeared to make a difference to the effect sizes. When externalizing behavior was observed directly (seven studies) or was indexed by a clinical diagnosis (six studies), the combined effect sizes were larger than in cases where a parent or teacher rated the level of externalizing behaviors. The group of studies based on observations of externalizing behavior was homogeneous. Its combined effect size of $d = 0.58$ was therefore an adequate estimate of the average effect, which amounted to a quite strong association between attachment security and externalizing behavior. The way in which attachment security was assessed also made a significant difference. In particular, studies conducted with the AQS showed the largest effect sizes ($d = 0.70$), whereas studies using the SSP yielded the lowest combined effect size ($d = 0.18$), though this latter effect was nevertheless significant. Because the different attachment measures are typically conducted at different ages, we also conducted a meta-regression analysis with age at the assessment of attachment as a predictor. As expected, the regression was significant (slope = .006, $p = .01$). Notably though, the effect of age was not significant within the SSP, AQS, or Cassidy and Marvin studies, suggesting that age and attachment measure were confounded. Furthermore, the regression with age did not appear to result solely from the larger effects associated with the AQS, as the same regression was significant when AQS studies were excluded (slope = .005, *p* = .03). It is also noteworthy that the overall effect size for security did not vary as a function of whether or not disorganization had been coded (*Q* = 0.10, *p* = .78). Note that this latter analysis excludes AQS studies by definition, as this procedure does not yield a score or classification for disorganized attachment.

Because the SSP is considered to be the gold standard for measuring attachment security, and the set of SSP studies was sufficiently large (*k* = 43), we decided to conduct complementary analyses for secure versus insecure attachment on this subset (see Table 2). In this homogeneous set of studies, we did not find a difference in combined effect size between clinical and nonclinical samples. The difference between samples with only girls versus only boys or mixed samples, however, was significant again, with all-female samples showing no association between attachment security and externalizing problems (*d* = −0.06). Also, we confirmed the larger combined effect size for those SSP studies that included observational measures of externalizing (*d* = 0.61) compared to the other types of assessment.

Interestingly, among the SSP studies the age at which the assessment of externalizing behavior was taken yielded a significant regression weight, with a slope of .002 (*p* = .02), indicating that the association between attachment security and externalizing became stronger with age. Because the SSPs are usually conducted within a small age window of 12–18 months, this significant slope suggested that the prediction of externalizing from attachment security was better with later—not earlier—assessments of externalizing, which is surprising given the longer interval between assessments in these studies.

Is Insecure-Avoidant Attachment Associated With More Externalizing Problems?

In 34 studies involving *N* = 3,675 participants, the insecure-avoidant attachment classifications were differentiated from the other classifications, and in this subset of studies the combined effect size was only *d* = 0.12, which was significant but small in magnitude (see Table 3). Insecure-avoidantly attached children displayed somewhat more externalizing behaviors than comparisons. With one study trimmed and filled the resulting significant effect size was *d* = 0.11, but the fail-safe number of studies needed to bring the effect down below significance was only 24. Because this number is below the Rosenthal (1991) criterion of 5*k* + 10, this outcome should be considered with caution. No significant moderator effects were found. The effect did not vary according to whether or not disorganization had been coded. Within the SSP subset of 25 studies, the combined effect size for insecure-avoidant attachment was *d* = 0.13, and the moderator analyses on this subset converged with the analyses on the total set of studies for avoidance (see Table 3 for those contrasts that could be tested).

Is Insecure-Resistant Attachment Associated With More Externalizing Problems?

In 35 studies involving *N* = 3,568 participants, the insecure-resistant attachment classifications were differentiated from the other classifications. In this subset of studies, the combined effect size was not significant, *d* = 0.11 (see Table 4), and no significant moderator effects were found. Within the SSP subset of 24 studies, the combined effect size for insecure-resistant attachment was only *d* = 0.05, and the contrasts on this subset showed no significant moderators (see Table 4 for those contrasts that could be tested).

Is Disorganized Attachment Associated With More Externalizing Problems?

In 34 studies including *N* = 3,778 participants, a significant combined effect size of *d* = 0.34 (see Table 5) was found for the association between disorganized attachment and externalizing behavior. As expected, disorganized attachment was associated with a higher risk for externalizing behavior later in childhood. However, eight studies had to be trimmed and filled, with a recomputed significant combined effect size of *d* = 0.18 (95% CI: 0.01, 0.34). The fail-safe number amounted to *k* = 407, which was above the Rosenthal criterion, suggesting that the file-drawer problem was not responsible for the association found in the current set of studies.

The set of studies was heterogeneous, and moderator analyses showed that gender was a significant moderator (see Table 5). Remarkably, in the samples with females only the association between disorganized attachment and externalizing behaviors was significantly different from the samples with only boys or with mixed gender, and in fact, the relation was negative; that is, disorganized attachment was associated with less externalizing behavior. However, the combined effect size of the six female samples (*N* = 702) was modest, *d* = −0.20. Clinical status, SES, or type of assessment (for both SSP and externalizing behavior) did not appear to moderate the association

Table 4
Resistant Attachment and Externalizing Behavior: Total Set and Set of SSP Studies

	k	N	d	Confidence interval 95%	Homogeneity Q	Contrast Q^a	Contrast p
Total set	35	3,568	0.11	−0.04, 0.26	66.32**		
Clinical						1.36	.24
Nonclinical	29	3,165	0.06	−0.10, 0.23	48.76**		
Clinical	6	403	0.30	−0.05, 0.64	14.27*		
Gender						0.11	.95
Boys	9	883	0.16	−0.02, 0.34	7.47		
Girls	10	809	0.16	−0.20, 0.52	22.88**		
Mixed	16	1,876	0.10	−0.12, 0.32	35.45**		
SES							
Low	3	857	0.04	−0.17, 0.24	0.66		
Middle/high	32	2,711	0.11	−0.06, 0.28	65.17**		
Measure						0.97^a	.33
SSP	24	2,910	0.05	−0.13, 0.24	35.54*		
CassMarvin	9	559	0.25*	0.05, 0.45	12.90		
Other	2	99	0.39	−0.23, 1.01	13.24**		
Observer ext							
Mother	27	2,831	0.08	−0.12, 0.28	54.62**		
Teacher	3	334	0.21	−0.29, 0.71	7.14*		
Observed	1	48	−0.07	−0.66, 0.52			
Other	3	250	0.24	−0.02, 0.49	1.03		
Combined	1	105	0.10	−0.29, 0.49			
SSP studies							
All SSP studies	24	2,910	0.05	−0.13, 0.24	35.54*		
Gender						0.56	.75
Boys	5	637	0.11	−0.14, 0.36	4.17		
Girls	7	766	0.17	−0.19, 0.53	16.59*		
Mixed	12	1,507	0.01	−0.13, 0.14	14.03		
SES							
Low	3	857	0.04	−0.17, 0.24	0.66		
Middle/high	21	2,053	0.05	−0.14, 0.23	34.87*		

Note. SES = socioeconomic status; SSP = Strange Situation procedure; CassMarvin = Cassidy and Marvin/MacArthur Preschool Attachment Coding System.
[a]Subgroups with $k < 4$ excluded from contrast.
*$p < .05$. **$p < .01$.

between disorganized attachment and externalizing behavior.

In the set of 24 SSP studies with $N = 3,161$ participants, we found a significant combined effect size of $d = 0.27$, which was similar to the effect size computed for the total set. No other significant moderators were found in this set (see Table 5 for those contrasts that could be tested).

Additional Analyses

The core set of studies on externalizing included assessments of externalizing as well as of aggressive behavior in the case of studies that did not present data on externalizing problems. In order to examine whether the more focused aggression studies would result in higher effect sizes, we decided to conduct two sets of meta-analyses, one for studies with data on externalizing and one for studies presenting aggression data. These two sets of studies partially overlapped (as some studies reported both), and it was therefore impossible to compare directly effect sizes across these sets (see the Method section). We computed 85% CIs for the point estimates of the combined effect sizes and compared these intervals across the two sets of studies: Nonoverlapping 85% CIs indicated a significant difference in combined effect sizes (Bakermans-Kranenburg et al., 2003). For the association between secure versus insecure attachment and externalizing behavior ($k = 65$), a combined effect size of $d = 0.32$ ($p < .01$; 85% CI: 0.26, 0.38) was found. The comparable combined effect size for the aggression outcomes ($k = 32$) was $d = 0.24$ ($p < .01$;

Table 5

Disorganized Attachment and Externalizing Behavior: Total Set and Set of SSP Studies

	k	N	d	Confidence interval 95%	Homogeneity Q	Contrast Q[a]	Contrast p
Total set	34	3,778	0.34**	0.18, 0.50	99.66**		
Clinical						0.75	.39
Nonclinical	27	3,184	0.30**	0.12, 0.49	83.00**		
Clinical	7	594	0.43**	0.26, 0.61	11.91		
Gender						6.58	.04
Boys	8	839	0.35*	0.03, 0.66	20.84**		
Girls	6	702	−0.20*	−0.39, −0.01	5.62		
Mixed	20	2,237	0.44**	0.26, 0.61	42.82**		
SES						0.42	.52
Low	9	1,266	0.44**	0.28, 0.60	13.42		
Middle/high	25	2,512	0.31**	0.12, 0.50	80.22**		
Measure						1.32[a]	.25
SSP	24	3,161	0.27**	0.10, 0.45	69.00**		
CassMarvin[b]	9	573	0.50**	0.16, 0.84	19.41*		
Other	1	44	1.10**	0.41, 1.79			
Observer ext						5.78[a]	.06
Mother	24	2,758	0.20*	0.02, 0.39	52.94**		
Teacher	4	415	0.48*	0.06, 0.91	11.73**		
Observed	0	0					
Other	4	345	0.62**	0.39, 0.84	5.32		
Combined	2	260	0.61*	0.08, 1.13	7.32**		
SSP studies							
All SSP studies	24	3,161	0.27**	0.10, 0.45	69.00**		
Clinical						0.08	.77
Nonclinical	20	2,817	0.29**	0.09, 0.48	67.25**		
Clinical	4	344	0.26*	0.03, 0.38	1.55		
Gender						10.99[a]	< .01
Boys	3	568	0.12	−0.12, 0.36	5.17		
Girls	4	669	−0.24*	−0.44, −0.05	1.84		
Mixed	17	1,924	0.39**	0.22, 0.56	31.38**		
SES						2.15	.14
Low	9	1,266	0.44**	0.28, 0.60	13.42		
Middle/high	15	1,895	0.17	−0.03, 0.38	43.81**		

Note. SES = socioeconomic status; SSP = Strange Situation procedure; CassMarvin = Cassidy and Marvin/MacArthur Preschool Attachment Coding System.
[a] Subgroup with $k < 4$ excluded from contrast. [b] Insecure-other included in disorganized.
*$p < .05$. **$p < .01$.

85% CI: 0.13, 0.35). For the association between avoidant versus nonavoidant attachments and externalizing ($k = 33$), the combined effects size was $d = 0.12$ ($p < .01$; 85% CI: 0.06, 0.21), and for the aggression outcomes ($k = 20$) the combined effect size amounted to $d = 0.28$ ($p < .05$; 85% CI: 0.08, 0.49). For the association between resistant versus nonresistant attachments and externalizing ($k = 34$), the combined effects size was $d = 0.10$ ($p > .05$; 85% CI: −0.01, 0.21), and for the aggression outcomes ($k = 22$) the combined effect size was $d = 0.05$ ($p > .05$; 85% CI: −0.10, 0.21). Finally, for the association between disorganized attachments and externalizing ($k = 33$), the combined effect size was $d = 0.37$ ($p < .01$; 85% CI: 0.25, 0.49), whereas for the aggression outcomes ($k = 16$) this combined effect size was $d = 0.12$ ($p > .05$; 85% CI: −0.06, 0.30). The 85% CIs did overlap for each of these comparisons, indicating that study outcomes with externalizing behavior or with aggression did not result in significantly different effect sizes.

Because the various comparisons (secure vs. insecure, avoidant vs. nonavoidant, and disorganized vs. nondisorganized) were based on varying numbers of studies and participants, we also selected a core set of 19 studies that provided data on all four comparisons. For the association between attachment security and externalizing, the combined effect size was $d = 0.27$ ($p < .01$; 85% CI: 0.15, 0.39). For avoidance, this combined effect size amounted to $d = 0.06$ ($p > .05$; 85% CI: −0.01, 0.13), for resistance it was $d = 0.11$ ($p < .05$; 85% CI: 0.03, 0.19), and for disorganization the combined effect size was $d = 0.29$ ($p < .01$; 85% CI: 0.13, 0.45). The 85% CIs of the combined effect sizes for attachment security and disorganization did not overlap with the 85% CI of the combined effect size for avoidance, but for the other comparisons the 85% CIs overlapped. Resistant attachment did not show a different combined effect size compared to all other attachment classifications. Insecure and disorganized attachments thus implicate a higher risk for externalizing behavior than avoidant attachment. We also compared each of the nonsecure classifications with the secure classification. In a core set of 18 studies that provided data on all three comparisons, the combined effect size for the association between avoidance versus security and externalizing was $d = 0.12$ ($p < .05$; 85% CI: 0.04, 0.20). For resistant versus secure attachment, it was $d = 0.19$ ($p < .01$; 85% CI: 0.11, 0.28), and for disorganized versus secure attachment, the combined effect size was $d = 0.27$ ($p < .01$; 85% CI: 0.13, 0.41). For all three comparisons, the 85% CIs overlapped, implying that the effects of the various types of insecurity showed similar associations with externalizing behavior.

Discussion

Since Bowlby's earliest work on attachment and separation, there have been persistent suggestions in the literature that attachment insecurity may play an important role in the development of aggression and antisocial behavior (Bowlby, 1944; Lewis et al., 1984; Lyons Ruth et al., 1993; Renken et al., 1989; van IJzendoorn, 1997). The body of research that subsequently tested this association is impressive in its sheer size. However, despite an extensive accumulation of data, a clear view on the empirical standing of this important hypothesis has been elusive. Apparently contradictory findings, nonreplications, and a diversity of study designs and sample sizes have created a body of work that is difficult to integrate coherently in narrative reviews.

The central question we thus posed in this meta-analysis was whether attachment insecurity was associated with externalizing behaviors across all the studies conducted to date. The results showed quite clearly that the answer to this question is a firm yes. Drawing from data on nearly 6,000 children tested in standardized observational assessments of mother–child attachment security, the average effect size for the contrast between secure and insecure children was $d = 0.31$ (95% CI: 0.23, 0.40). Over 1,700 studies of average sample size with null results would need to be added to the database to reduce this effect to nonsignificance. For clinical samples, the average effect size amounted to $d = 0.49$ (95% CI: 0.32, 0.66). On the face of it, these robust findings lend direct support to the notion that attachment plays a significant role in the evolution of children's behavior problems, for typically developing children as well as for clinical groups.

It should be noted that for meta-analytic results the evaluation of combined effect sizes in terms of absolute magnitude is problematic (McCartney & Rosenthal, 2000). It is arguably more meaningful to consider the global association between insecurity and externalizing problems in the context of other studies examining similar phenomena and employing similar methodologies (McCartney & Rosenthal, 2000). In that respect, the combined effect size of $d = 0.31$ reported here is of similar magnitude to meta-analytic results concerning the association between aggression and hostile attributional biases ($d = 0.35$; see Orobio de Castro, Veerman, Koops, Bosch, & Monshouwer, 2002) or resting heart rate ($d = -0.38$, Lorber, 2004) and substantially higher than that between aggression and basal cortisol ($d = -0.10$; see Alink et al., 2008). The combined effect found in this analysis gains even greater significance in light of the fact that a majority of the studies included in the analysis were longitudinal investigations, where independent and dependent variables were measured often several years apart (with an intervening period of 25 months on average).

Three perhaps obvious additional points about the effects should be made at the outset. First, there were so few outcome studies that examined father–child attachment security that we were unable to include them in this meta-analysis. There is clearly an urgent need for further research into the contribution of father–child attachment security and insecurity to children's development. Second, the effects reported in this meta-analysis reflect statistical association, not causation and—although they provide evidence relevant to efforts to determine causality and its mechanisms—alone they are mute on this issue. Third, the effects are uncorrected for the influence of relevant third variables. Controls for these could reduce or increase the magnitude of the global meta-analytic effects. Furthermore, the effects represent an average across a potentially large number of moderating relationships that could amplify or attenuate the association between attachment and externalizing problems (e.g., Belsky & Fearon, 2002; NICHD Early Child Care Research Network, 2006). Related to this point, the association for children from nonclinical populations was highly heterogeneous, indicating large between-study differences in effect size, which could have resulted from a number of methodological factors, such as the composition of the populations studied, the length of the follow-up period or the measurement strategies that were adopted. Indeed, the larger effect found in clinical samples ($d = 0.49$) illustrates this point, and at the same time demonstrates the significance of attachment for clinical groups.

Within clear constraints, the analyses reported in this study were able to consider the role of a number of other potentially important moderators that have been highlighted in the literature. Central among these were SES (with low-risk samples anticipated to yield larger effects), gender (insecure boys expected to show more behavior problems), and age of outcome (with effects expected to persist over time). In contrast to our expectations, the magnitude of the association between attachment and behavior problems was relatively consistent across high- and low-SES samples ($d = 0.25$ vs. $d = 0.31$, respectively). SES represents a rather blunt approximation of a number of important proximal psychosocial risk processes that may be more clearly implicated in the association between attachment and children's externalizing problems. Nevertheless, SES does account for a considerable portion of the variance in externalizing problems in childhood (Bradley & Corwyn, 2002), and the absence of moderation in this meta-analysis is a potentially important result. The finding certainly suggests that attachment insecurity is associated with higher levels of behavior problems even in apparently low-risk psychosocial circumstances. However, possible effects of SES should not be ruled out, as a broad division of samples into low versus high or middle class is clearly limited in precision and in several cases the reviewed studies involved mixtures of differing levels of socioeconomic disadvantage that were not captured in our coding scheme.

Turning to gender, the results were consistent with our expectations and the findings of some early longitudinal studies (Lewis et al., 1984; Renken et al., 1989), in that attachment was more strongly associated with externalizing behavior problems in samples of boys than in samples of girls ($d = 0.35$ vs. $d = -0.03$, respectively). The set of studies on girls was homogeneous; therefore, the absence of a significant association was not due to one or a few outlying outcomes in the lower range. Furthermore, within the mixed sample (boys and girls), a metaregression showed a trend for stronger effects in samples with comparatively more boys than girls. Given that the variation in gender composition within this subset reflected a very narrow range, this finding—when combined with those from the single-gender studies—provides quite compelling evidence of the elevated significance of attachment for behavior problems in boys. Nevertheless, the mixed samples, constituting the large majority of the studies, showed equally strong effect sizes as those with only boys, suggesting that the association between attachment and externalizing cannot be ascribed to boys only.

There are a number of plausible ways of interpreting this gender effect. First, on the face of it the effect may not be surprising, given the substantially higher risk of externalizing problems in boys (Loeber & Hay, 1997). However, this broad explanation subsumes several distinct possibilities that are worthy of attention. First, the lower rates of externalizing behavior problems in girls may impose a range restriction on the dependent variable, which in turn would attenuate the effect size (DeKlyen & Greenberg, 2008). On the other hand, the lower rates of externalizing problems may represent a manifestation of a differing set of etiological mechanisms in girls than boys, with attachment processes figuring more significantly in the developmental trajectories of boys. Certainly, some behavior genetic studies have suggested partially independent genetic and environmental contributions to externalizing symptoms (e.g., Vierikko, Pulkkinen, Kaprio, Viken, & Rose, 2003). Furthermore, a number of nongenetic studies have documented distinct risk factors for girls and boys (Cairns & Cairns, 1984; Cummings, Iannotti, & Zahn Waxler, 1989). A further important possibility to consider is that attachment (and indeed other risk factors) may contribute to a common latent process that has distinctive behavioral manifestations, which in turn are more commonly associated with boys than girls. Pertinent candidate examples might include the behavioral and contextual differences between physical and relational aggression (Crick & Grotpeter, 1995) or overt versus covert antisocial behavior (Loeber & Schmaling, 1985). To the extent that these behavioral processes are measured by different instruments, are evident in different contexts, or are weighted differently by different observers, we may expect differential associations by gender. A further important possibility to consider, which may not be independent of the possibilities described earlier, is that insecurity in girls is more associated with internalizing problems rather than externalizing problems (see DeKlyen & Greenberg, 2008).

In light of claims made by many in the field that attachment has persistent effects on future development, we also expected the association between attachment and behavior problems to be evident in studies with relatively long-term follow-up periods. Consistent with this expectation, we found that the age of the child at the point when externalizing problems were assessed was not significantly associated with the magnitude of study effect sizes. Even more intriguingly, among studies employing the SSP ($k = 43$ studies), applied within a small time window, effect sizes appeared to increase significantly over time.

In addition to these a priori hypotheses, method of measurement proved a significant factor that distinguished studies with relatively small and large effects in two distinct ways. First, different assessments of attachment appeared to be associated with reliable differences in the magnitude of the attachment-behavior problems association. Broadly speaking, the SSP produced smaller effect sizes than the other attachment assessments and the AQS produced comparatively larger effect sizes. Critically, however, it was not possible to disentangle entirely the role played by measurement type (e.g., SSP vs. other assessments) from the age at which the assessment was conducted (infancy or later), as these were essentially confounded. Notably, age was not a significant regressor within the SSP studies or within the non-SSP studies, so a substantial portion of the overall effect came exclusively from the difference between the SSP studies on the one hand and the other assessment types on the other. Although the AQS evidenced the strongest effect sizes by some margin ($d = 0.70$), the metaregression of age on effect size remained significant even when the AQS was removed from the analysis. Thus, while it was clear that security assessed by the AQS was rather more strongly associated with behavior problems than that derived from the SSP, later assessments of attachment beyond the SSP also appeared to yield stronger effect sizes. On that basis, it was not possible to determine whether the critical factor was some methodological or coding variation that distinguishes the SSP from the other attachment assessments, or some factor more closely related to development itself. It is certainly plausible that important developmental changes take place around the beginning of the 3rd year that amplify the link between attachment and externalizing problems.

The second critical measurement factor that emerged from this meta-analysis concerned the measurement of outcome. Studies that assessed externalizing behavior problems via direct observation identified reliably larger effect sizes ($d = 0.58$) than those that relied on questionnaires from parents or teachers ($d = 0.22$ and $d = 0.30$, respectively). This is a potentially important finding from a variety of perspectives. First, the vast majority of the studies that investigated externalizing behavior in relation to attachment reviewed here did not use any objective source of information concerning the outcome. Not only does this strategy limit the conclusions that can be drawn (see Kagan, 2007), it could also partially explain the mixed effects across attachment-outcome studies as a whole. It is noteworthy that the effect sizes cited earlier are apparently positively correlated with the degree of objectivity of the observer.

Nevertheless, although the objectivity of measurement is one potential account of the stronger effect sizes associated with direct observation, it could also be that direct observation reveals qualitatively distinct features of behavior or qualitatively distinct contexts in which behavior takes place. Notably, studies that used observational methods typically focused on the peer setting (e.g., Booth, Rose Krasnor, & Rubin, 1991; Matas et al., 1978; Suess, Grossmann, & Sroufe, 1992) and hence may have been tapping behavior more closely connected with social competence than those that relied on parent or teacher questionnaires. However, these effects are understood, the limited agreement between objective observers, teachers, and parents concerning children's behavior problems creates a strong imperative to use multiple sources of outcome data in future studies of attachment security and its sequelae. At the same time, more data sources per se will not solve this problem. What is arguably needed is a better understanding of the precise circumstances and mechanisms under which aggression and other antisocial acts may be triggered and the contribution that attachment processes make to this. It is hard to imagine that this degree of mechanistic specificity could be achieved without greater reliance on field studies that involve direct and extensive observations in a range of relevant social settings.

Arguments put forward in the literature, and results of an earlier meta-analysis, led us to expect that attachment disorganization would be a stronger predictor of externalizing behavior problems than attachment insecurity generally (van IJzendoorn et al., 1999). However, the findings of the current meta-analysis were only partially supportive of the special status sometimes accorded to disorganized attachment as a precursor of children's externalizing problems. The global effect size for attachment security (*d* = 0.31) was very similar to that for disorganization (*d* = 0.34). It should be noted that these effect sizes do not reflect the outcomes of the same set of studies so direct comparisons are difficult. However, in an analysis of 19 studies where data from all four attachment groups were available the results corresponded to the aforementioned global effects. The effects for security and disorganization were significant and of similar magnitude (*d* = 0.27 and *d* = 0.29, respectively), while the effects for avoidance and resistance were small and in the case of avoidance nonsignificant (*d* = 0.06 and *d* = 0.11, respectively). The 85% CIs for security and disorganization did not overlap with that for avoidance, suggesting significantly stronger effects for insecurity and disorganization than for avoidance. The effect for resistance did not differ from any of the other contrasts, in part because this effect showed marked heterogeneity. Furthermore, in the larger set of studies, resistant attachment was not significantly associated with externalizing problems and showed an effect size numerically similar to that for avoidance. When only studies that provided direct pairwise comparisons between groups were considered, secure children scored lower on measures of externalizing problems than disorganized (*d* = 0.27), avoidant (*d* = 0.12), and resistant (*d* = 0.19) children, and while the CIs for these effects overlapped, disorganized children demonstrated numerically larger effects than the other insecure categories.

Although the effect sizes for disorganization and security compare favorably with other prominent risk factors for externalizing behavior, they also leave room for the possibility that the association may be moderated by other factors. It is highly noteworthy that few if any of the studies reviewed in this study directly addressed questions of mechanism. Arguably, until mediating mechanisms are better understood, we may struggle to find the relevant moderators. Given that an association clearly exists, there is an obvious need for strong theory-driven studies that address mediating processes, particularly those drawing on methods derived from other fields. Some well-studied risk processes worthy of consideration (in addition to the traditional internal working models construct) include impulsivity, negativity emotionality, affect regulation, hostile attributional biases, and physiological hypo-arousal (Belsky, Fearon, & Bell, 2007; Dodge & Coie, 1987; Eisenberg et al., 2001; Raine, Venables, & Mednick, 1997). Risk factors such as these, situated at the biological, cognitive, or affective level may be considered proximal determinants of externalizing behavior, with the quality of the attachment relationship with a primary caregiver conceptualized as a more distal determinant. A crucial question in that context has long been, and remains, the extent to which longitudinal continuities in the effects of attachment represent the ongoing supportive function provided by the attachment relationship as opposed to the early effects of attachment experiences on the emergence of stable psychological structures, such as internal working models. It is notable that Belsky and Fearon's (2002) analysis of the NICHD data at age 1–3 years suggested that the effects of attachment tended to persist primarily when there was continuity in the quality of maternal care. Given the centrality of this issue and the cogent arguments made by several authors on this subject (e.g., Sameroff, 2000; Sroufe et al., 2005a, 2005b), it is unfortunate that few studies have attempted to address this. Related to this, few studies have directly considered the possibility that certain parenting characteristics may be common determinants of both attachment insecurity and externalizing behavior problems. Thus, the extent to which attachment processes per se can be shown to make a specific and causal contribution to children's externalizing problems, either independently of parenting or as a causal mediator of its effects, remains to be seen.

There is also a clear need for a better understanding of causation. In that regard, the absence of studies that have repeatedly assessed both attachment and outcome is a significant barrier. Longitudinal studies that employ cross-lagged panel designs could provide elegant tests of causation by focusing on temporal ordering. However, in order to do this there is an urgent need to establish robust measurement protocols that allow for meaningful repeated assessments of attachment and hence the documentation of change. Of course, causal hypotheses can also be powerfully addressed by intervention studies (Bakermans-Kranenburg, van IJzendoorn, Mesman, Alink, & Juffer, 2008) and in the future these may be critical for determining the role of attachment in children's behavior problems, as well as that of the causal mediators and moderators of its effects.

References

References with asterisks are included in the meta-analytic data set.

Ainsworth, M. S., Blehar, M. C., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the strange situation* (Vol. 18). Hillsdale, NJ: Erlbaum.

Alink, L. R. A., Van IJzendoorn, M. H., Bakermans-Kranenburg, M. J., Mesman, J., Juffer, F., & Koot, H. M. (2008). Cortisol and externalizing behavior in children and adolescents: Mixed meta-analytic evidence for the inverse relation of basal cortisol and cortisol reactivity with externalizing behavior. *Developmental Psychobiology, 50*, 427–450.

*Anan, R. M., & Barnett, D. (1999). Perceived social support mediates between prior attachment and subsequent adjustment: A study of urban African American children. *Developmental Psychology, 35*, 1210–1222.

*Aviezer, O., Sagi, A., Resnick, G., & Gini, M. (2002). School competence in young adolescence: Links to early attachment relationships beyond concurrent self-perceived competence and representations of relationships. *International Journal of Behavioral Development, 26*, 397–409.

*Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2006). Gene–environment interaction of the dopamine D4 receptor (DRD4) and observed maternal insensitivity predicting externalizing behavior in preschoolers. *Developmental Psychobiology, 48*, 406–409.

Bakermans-Kranenburg, M. J., van IJzendoorn, M. H., & Juffer, F. (2003). Less is more: Meta-analyses of sensitivity and attachment interventions in early childhood. *Psychological Bulletin, 129*, 195–215.

Bakermans-Kranenburg, M. J., van IJzendoorn, M. H., Mesman, J., Alink, L. R. A., & Juffer, F. (2008). Effects of an attachment-based intervention on daily cortisol moderated by DRD4: A randomized control trial on 1-3-year-olds screened for externalizing behavior. *Development & Psychopathology, 20*, 805–820.

Baldwin, J. M. (1895). *Mental development of the child and the race: Methods and processes*. New York: Macmillan.

*Bates, J. E., & Bayles, K. (1988). Attachment and the development of behavior problems. In J. Belsky & T. Nezworski (Eds.), *Clinical implications of attachment. Child psychology* (Vol. 7, pp. 253–299). Hillsdale, NJ: Erlbaum.

Bates, J., Maslin, C., & Frankel, K. (1985). Attachment security, mother-child interactions, and temperament as predictors of behavior problem ratings at age three years. In I. Bretherton & E. Waters (Eds.), Growing points in attachment theory and research. *Monographs of the Society for Research in Child Development, 50*(1–2, Serial No. 209), 167–193.

Belsky, J., & Fearon, R. M. P. (2002). Infant-mother attachment security, contextual risk, and early development: A moderational analysis. *Development and Psychopathology, 1*, 293–310.

Belsky, J., Fearon, R. M. P., & Bell, B. (2007). Parenting, attention and externalizing problems: Testing mediation longitudinally, repeatedly and reciprocally. *Journal of Child Psychology and Psychiatry, 48*, 1233–1242.

Belsky, J., & Isabella, R. (1988). Maternal, infant, and social-contextual determinants of attachment security. In J. Belsky & T. Nezworski (Eds.), *Clinical implications of attachment* (pp. 41–94). Hillsdale, NJ: Erlbaum.

Belsky, J., & Nezworski, T. (Eds.). (1988). *Clinical implications of attachment*. Hillsdale, NJ: Erlbaum.

*Booth, C. L., Rose Krasnor, L., McKinnon, J. A., & Rubin, K. H. (1994). Predicting social adjustment in middle childhood: The role of preschool attachment security and maternal style. *Social Development, 3*, 189–204.

*Booth, C. L., Rose Krasnor, L., & Rubin, K. H. (1991). Relating preschoolers' social competence and their mothers' parenting behaviors to early attachment security and high-risk status. *Journal of Social and Personal Relationships, 8*, 363–382.

Borenstein, M., Rothstein, D., & Cohen, J. (2005). Comprehensive meta-analysis: A computer program for research synthesis [Computer software]. Englewood Cliffs, NJ: Biostat.

Bowlby, J. (1944). Fourty-four juvenile theives: Their characters and home life. *International Journal of Psycho-Analysis, 25*, 19–52.

Bowlby, J. (1969). *Attachment and loss: Vol. 1. Attachment*. London: Hogarth Press and the Institute of Psycho-Analysis.

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology, 53*, 371–399.

Bretherton, I. (1995). Internal working models: Cognitive and affective aspects of attachment representations. In D. Cicchetti & S. Toth (Eds.), *4th Rochester Symposium on Developmental Psychopathology on "Emotion, Cognition, and Representation"* (pp. 231–259). Hillsdale, NJ: Erlbaum.

Bretherton, I. (1997). Bowlby's legacy to developmental psychology. *Child Psychiatry and Human Development, 28*, 33–43.

*Burgess, K. B., Marshall, P. J., Rubin, K. H., & Fox, N. A. (2003). Infant attachment and temperament as predictors of subsequent externalizing problems and cardiac physiology. *Journal of Child Psychology and Psychiatry, 44*, 819–831.

Cairns, R. B., & Cairns, B. D. (1984). Predicting aggressive patterns in girls and boys: A developmental study. *Aggressive Behavior, 10*, 227–242.

*Carlson, E. A. (1998). A prospective longitudinal study of attachment disorganization/disorientation. *Child Development, 69*, 1107–1128.

Cassidy, J. (1988). Child-mother attachment and the self in six-year-olds. *Child Development, 59*, 121–134.

Cassidy, J. (1994). Emotion regulation: Influences of attachment relationships. In N. A. Fox (Ed.), *The development of attachment regulation* (pp. 228–249). *Monograph of the Society for Research in Child Development, 59* (Serial No. 240).

Cassidy, J., Marvin, R. S., & The MacArthur Working Group on Attachment. (1989). *Attachment organization* *in three and four year olds: Coding guidelines.* Unpublished scoring manual, University of Virginia, Charlottesville.

*Cicchetti, D., Rogosch, F. A., & Toth, S. L. (1998). Maternal depressive disorder and contextual risk: Contributions to the development of attachment insecurity and behavior problems in toddlerhood. *Development and Psychopathology, 10*, 283–300.

*Cohn, D. A. (1990). Child-mother attachment of six-year-olds and social competence at school. *Child Development, 61*, 152–162.

Crick, N. R., & Grotpeter, J. K. (1995). Relational aggression, gender, and social-psychological adjustment. *Child Development, 66*, 710–722.

Cummings, E. M., Iannotti, R. J., & Zahn Waxler, C. (1989). Aggression between peers in early childhood: Individual continuity and developmental change. *Child Development, 60*, 887–895.

DeKlyen, M., & Greenberg, M. T. (2008). Attachment and psychopathology in childhood. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of attachment: Theory, research and clinical applications* (2nd ed., pp. 637–665). New York: Guilford.

*DeMulder, E. K., Denham, S., Schmidt, M., & Mitchell, J. (2000). Q-sort assessment of attachment security during the preschool years: Links from home to school. *Developmental Psychology, 36*, 274–282.

*DeVito, C., & Hopkins, J. (2001). Attachment, parenting, and marital dissatisfaction as predictors of disruptive behavior in preschoolers. *Development and Psychopathology, 13*, 215–231.

Dodge, K. A., & Coie, J. D. (1987). Social-information-processing factors in reactive and proactive aggression in children's peer groups. *Journal of Personality and Social Psychology, 53*, 1146–1158.

Duval, S., & Tweedie, R. (2000a). A nonparametric "trim and fill" method for accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95*, 89–98.

Duval, S., & Tweedie, R. (2000b). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*, 455–463.

*Edwards, E. P., Eiden, R. D., & Leonard, K. E. (2006). Behavior problems in 18- to 36-month-old children of alcoholic fathers: Secure mother-infant attachment as a protective factor. *Development and Psychopathology, 18*, 395–407.

*Egeland, B., & Heister, M. (1995). The long-term consequences of infant day-care and mother-infant attachment. *Child Development, 66*, 474–485.

Eisenberg, N., Cumberland, A., Spinrad, T. L., Fabes, R. A., Shepard, S. A., Reiser, M., et al. (2001). The relations of regulation and emotionality to children's externalizing and internalizing problem behavior. *Child Development, 72*, 1112–1134.

Erickson, M. F., Sroufe, L. A., & Egeland, B. (1985). The relationship between quality of attachment and behavior problems in preschool in a high-risk sample. In I. Bretherton & E. Waters (Eds.), Growing points of attachment theory and research. *Monographs of the Society for Research in Child Development, 50*(1–2, Serial No. 209), 147–166.

*Fagot, B. I., & Leve, L. D. (1998). Teacher ratings of externalizing behavior at school entry for boys and girls: Similar early predictors and different correlates. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 39*, 555–566.

Fonagy, P. (2004). Early-life trauma and the psychogenesis and prevention of violence. *Annals of the New York Academy of Sciences, 1036*, 181–200.

Freud, S. (1908). *Ueber infantile Sexualtheorien.* Wissenschaft: Studienausgabe Band V. Fischer.

Gilbody, S. M., Song, F., Eastwood, A. J., & Sutton, A. J. (2000). The causes, consequences and detection of publication bias in psychiatry. *ACTA Psychiatrica Scandinavica, 102*, 241–249.

*Gilliom, M., Shaw, D. S., Beck, J. E., Schonberg, M. A., & Lukon, J. L. (2002). Anger regulation in disadvantaged preschool boys: Strategies, antecedents, and the development of self-control. *Developmental Psychology, 38*, 222–235.

Goldberg, S. (1997). Attachment and childhood behavior problems in normal, at-risk and clinical samples. In L. Atkinson & K. J. Zucker (Eds.), *Attachment and psychopathology* (Vol. 3, pp. 171–195). New York: Guilford.

*Goldberg, S., Corter, C., Lojkasek, M., & Minde, K. (1990). Prediction of behavior problems in 4-year-olds born prematurely. *Development and Psychopathology, 2*, 15–30.

*Goldberg, S., Gotowiec, A., & Simmons, R. J. (1995). Infant-mother attachment and behavior problems in healthy and chronically ill preschoolers. *Development and Psychopathology, 7*, 267–282.

Goldstein, H., & Healy, M. J. R. (1995). The graphical presentation of a collection of means. *Journal of the Royal Statistical Society, Series A (Statistics in Society), 158*, 175–177.

*Greenberg, M. T., Speltz, M. L., Deklyen, M., & Endriga, M. C. (1991). Attachment security in preschoolers with and without externalizing behavior problems: A replication. *Development and Psychopathology, 3*, 413–430.

Guttmann-Steinmetz, S., & Crowell, J. A. (2006). Attachment and externalizing disorders: A developmental psychopathology perspective. *Journal of the American Academy of Child and Adolescent Psychiatry, 45*, 440–451.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* San Diego, CA: Academic Press.

*Howes, C., Matheson, C. C., & Hamilton, C. E. (1994). Maternal, teacher, and child care history correlates of children's relationships with peers. *Child Development, 65*, 264–273.

*Hubbs Tait, L., Osofsky, J. D., Hann, D. M., & McDonald Culp, A. (1994). Predicting behavior problems and social competence in children of adolescent mothers. *Family Relations: Interdisciplinary Journal of Applied Family Studies, 43*, 439–446.

James, W. (1890). *Principles of psychology*. New York: Henry Holt.

Kagan, J. (2007). A trio of concerns. *Perspectives on Psychological Science*, *2*, 361–376.

*Keller, T. E., Spieker, S. J., & Gilchrist, L. (2005). Patterns of risk and trajectories of preschool problem behaviors: A person-oriented analysis of attachment in context. *Development and Psychopathology*, *17*, 349–384.

*Klein Velderman, M., Bakermans-Kranenburg, M. J., Juffer, F., & van IJzendoorn, M. H. (2006). Effects of attachment-based interventions on maternal sensitivity and infant attachment: Differential susceptibility of highly reactive infants. *Journal of Family Psychology*, *20*, 266–274.

Kobak, R., Cassidy, J., Lyons Ruth, K., & Ziv, Y. (2005). Attachment, stress and psychopathology: A developmental pathways model. In D. Cicchetti & J. D. Cohen (Eds.), *Developmental psychopathology, Vol. I: Theory and method* (2nd ed., pp. 333–369). Hoboken, NJ: Wiley.

Kochanska, G. (1997). Mutually responsive orientation between mothers and their young children: Implications for early socialization. *Child Development*, *68*, 94–112.

Lamb, M. E., Thompson, R. A., Gardner, W. P., Charnov, E. L., & Estes, D. (1984). Security of infantile attachment as assessed in the "Strange Situation": Its study and biological interpretation. *Behavoral and Brain Sciences*, *7*, 127–172.

*Lewis, M., Feiring, C., McGuffog, C., & Jaskir, J. (1984). Predicting psychopathology in six-year-olds from early social relations. *Child Development*, *55*, 123–136.

*Lieberman, A. F., Weston, D. R., & Pawl, J. H. (1991). Preventive intervention and outcome with anxiously attached dyads. *Child Development*, *62*, 199–209.

Liotti, G. (1992). Disorganized/disoriented attachment in the etiology of the dissociative disorders. *Dissociation*, *5*, 196–204.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis. Applied social research methods series* (Vol. 49). Thousand Oaks, CA: Sage.

Loeber, R., & Hay, D. (1997). Key issues in the development of aggression and violence from childhood to early adulthood. *Annual Review of Psychology*, *48*, 371–410.

Loeber, R., & Schmaling, K. B. (1985). Empirical evidence for overt and covert patterns of antisocial conduct problems: A meta-analysis. *Journal of Abnormal Child Psychology*, *13*, 337–353.

Lorber, M. F. (2004). Psychophysiology of aggression, psychopathy, and conduct problems: A meta-analysis. *Psychological Bulletin*, *130*, 531–552.

*Lyons Ruth, K., Alpern, L., & Repacholi, B. (1993). Disorganized infant attachment classification and maternal psychosocial problems as predictors of hostile-aggressive behavior in the preschool classroom. *Child Development*, *64*, 572–585.

Lyons Ruth, K., Zeanah, C. H., & Benoit, D. (2003). Disorder and risk for disorder during infancy and toddlerhood. In E. J. Mash & R. A. Barkley (Eds.), *Child psychopathology* (2nd ed., pp. 589–631). New York: Guilford.

*Madigan, S., Moran, G., Schuengel, C., Pederson, D. R., & Otten, R. (2007). Unresolved maternal attachment representations, disrupted maternal behavior, and disorganized attachment in infancy: Links to toddler behavior problems. *Journal of Child Psychology and Psychiatry*, *48*, 1042–1050.

Main, M., & Cassidy, J. (1988). Categories of response to reunion with the parent at age 6: Predictable from infant attachment classifications and stable over a 1-month period. *Developmental Psychology*, *24*, 415–426.

Main, M., Kaplan, N., & Cassidy, J. (1985). Security in infancy, childhood and adulthood: A move to the level of representation. In I. Bretherton & E. Waters (Eds.), Growing points of attachment theory and research. *Monographs of the Society for Research in Child Development*, *50*(1–2, Serial No. 209), 66–104.

Main, M., & Morgan, H. (1996). Disorganization and disorientation in infant strange situation behavior: Phenotypic resemblance to dissociative states. In L. K. Michelson & W. J. Ray (Eds.), *Handbook of dissociation: Theoretical, empirical, and clinical perspectives* (pp. 107–138). New York: Plenum.

Main, M., & Solomon, J. (1986). Discovery of an insecure-disorganized/disoriented attachment pattern. In T. B. Brazelton & M. W. Yogman (Eds.), *Affective development in infancy* (pp. 95–124). Norwood: Ablex.

Main, M., & Solomon, J. (1990). Procedures for identifying infants as disorganized/disoriented during the Ainsworth Strange Situation. In M. T. Greenberg & D. Cicchetti (Eds.), *Attachment in the preschool years: Theory, research, and intervention. The John D. and Catherine T. MacArthur Foundation series on mental health and development* (Vol. 19, pp. 121–160). Chicago: University of Chicago Press.

*Marchand, J. F., & Hock, E. (1998). The relation of problem behaviors in preschool children to depressive symptoms in mothers and fathers. *Journal of Genetic Psychology*, *159*, 353–366.

*Matas, L., Arend, R. A., & Sroufe, L. A. (1978). Continuity of adaptation in the second year: The relationship between quality of attachment and later competent functioning. *Child Development*, *49*, 547–556.

McCartney, K., & Rosenthal, R. (2000). Effect size, practical importance, and social policy for children. *Child Development*, *71*, 173–180.

Moss, E., Bureau, J. F., Cyr, C., Mongeau, C., & St Laurent, D. (2004). Correlates of attachment at age 3: Construct validity of the preschool attachment classification system. *Developmental Psychology*, *40*, 323–334.

*Moss, E., Cyr, C., & Dubois-Comtois, K. (2004). Attachment at early school age and developmental risk: Examining family contexts and behavior problems of controlling-caregiving, controlling-punitive, and behaviorally disorganized children. *Developmental Psychology*, *40*, 519–532.

Moss, E., Parent, S., Gosselin, C., & Rousseau, D. (1996). Attachment and teacher-reported behavior problems during the preschool and early school-age period. *Development and Psychopathology*, *8*, 511–525.

Moss, E., Smolla, N., Cyr, C., Dubois Comtois, K., Mazzarello, T., & Berthiaume, C. (2006). Attachment and behavior problems in middle childhood as reported by adult and child informants. *Development and Psychopathology*, *18*, 425–444.

Mullen, B. (1989). *Advanced BASIC meta-analysis*. Hillsdale, NJ: Erlbaum.

Mullen, B., & Rosenthal, R. (1985). *BASIC meta-analysis*. Hillsdale, NJ: Erlbaum.

*Munson, J. A., McMahon, R. J., & Spieker, S. J. (2001). Structure and variability in the developmental trajectory of children's externalizing problems: Impact of infant attachment, maternal depressive symptomatology, and child sex. *Development and Psychopathology*, *13*, 277–296.

National Institute of Child Health and Human Development (NICHD) Early Child Care Research Network. (2006). Infant-mother attachment classification: Risk and protection in relation to changing maternal caregiving quality. *Developmental Psychology*, *42*, 38–58.

Orobio de Castro, B., Veerman, J. W., Koops, W., Bosch, J. D., & Monshouwer, H. J. (2002). Hostile attribution of intent and aggressive behavior: A meta-analysis. *Child Development*, *73*, 916–934.

*Perez Corres, A. (2006). *Mother's attachment representations as a contextual determinant of children's mental health. A sixteen-year longitudinal study*. Unpublished doctoral dissertation, University of London, United Kingdom.

*Pierrehumbert, B., Miljkovitch, R., Plancherel, B., Halfon, O., & Ansermet, F. (2000). Attachment and temperament in early childhood; Implications for later behavior problems. *Infant and Child Development*, *9*, 17–32.

*Radke Yarrow, M., McCann, K., DeMulder, E., Belmont, B., et al. (1995). Attachment in the context of high-risk conditions. *Development and Psychopathology*, *7*, 247–265.

Raine, A., Venables, P. H., & Mednick, S. A. (1997). Low resting heart rate at age 3 years predisposes to aggression at age 11 years: Evidence from the Mauritius Child Health Project. *Journal of the American Academy of Child and Adolescent Psychiatry*, *36*, 1457–1464.

Renken, B., Egeland, B., Marvinney, D., Mangelsdorf, S., et al. (1989). Early childhood antecedents of aggression and passive-withdrawal in early elementary school. *Journal of Personality*, *57*, 257–281.

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park, CA: Sage.

*Rothbaum, F., Rosen, K. S., Pott, M., & Beatty, M. (1995). Early parent-child relationships and later problem behavior: A longitudinal study. *Merrill-Palmer Quarterly*, *41*, 133–151.

Rutter, M. (1995). Clinical implications of attachment concepts: Retrospect and prospect. *Journal of Child Psychology and Psychiatry*, *36*, 549–572.

Sameroff, A. J. (2000). Developmental systems and psychopathology. *Development and Psychopathology*, *12*, 297–312.

*Schmidt, M. E., Demulder, E. K., & Denham, S. (2002). Kindergarten social-emotional competence: Developmental predictors and psychosocial implications. *Early Child Development and Care*, *172*, 451–462.

Schneider-Rosen, K., & Rothbaum, F. (1993). Quality of parental caregiving and security of attachment. *Developmental Psychology*, *29*, 358–367.

*Seifer, R., LaGasse, L. L., Lester, B., Bauer, C. R., Shankaran, S., Bada, H. S., et al. (2004). Attachment status in children prenatally exposed to cocaine and other substances. *Child Development*, *75*, 850–868.

*Shaw, D. S., Owens, E. B., Vondra, J. I., Keenan, K., & Winslow, E. B. (1996). Early risk factors and pathways in the development of early disruptive behavior problems. *Development and Psychopathology*, *8*, 679–699.

Shaw, D. S., & Vondra, J. I. (1995). Infant attachment security and maternal predictors of early behavior problems: A longitudinal study of low-income families. *Journal of Abnormal Child Psychology*, *23*, 335–357.

*Smeekens, S., Riksen Walraven, J. M., & van Bakel, H. (2007). Multiple determinants of externalizing behavior in 5-year-olds: A longitudinal model. *Journal of Abnormal Child Psychology*, *35*, 347–361.

Solomon, J., & George, C. (1999). The place of disorganization in attachment theory: Linking classic observations with contemporary findings. In J. Solomon & C. George (Eds.), *Attachment disorganization* (Vol. 23, pp. 3–32). New York: Guilford.

*Solomon, J., George, C., & De Jong, A. (1995). Children classified as controlling at age six: Evidence of disorganized representational strategies and aggression at home and at school. *Development and Psychopathology*, *7*, 447–463.

*Speltz, M. L., DeKlyen, M., & Greenberg, M. T. (1999). Attachment in boys with early onset conduct problems. *Development and Psychopathology*, *11*, 269–285.

*Speltz, M. L., Greenberg, M. T., & DeKlyen, M. (1990). Attachment in preschoolers with disruptive behavior: A comparison of clinic-referred and non-problem children. *Development and Psychopathology*, *2*, 31–46.

Sroufe, L. A. (1983). *Infant-caregiver attachment and patterns of adaptation in preschool: The roots of maladaption and competence* (Vol. 16). Hillsdale, NJ: Erlbaum.

Sroufe, L. A., Egeland, B., Carlson, E., & Collins, W. A. (2005a). *The development of the person: The Minnesota study of risk and adaptation from birth to adulthood*. New York: Guilford.

Sroufe, L. A., Egeland, B., Carlson, E., & Collins, W. A. (2005b). Placing early attachment experiences in developmental context: The Minnesota longitudinal study. In K. E. Grossmann, K. Grossmann, & E. Waters (Eds.), *Attachment from infancy to adulthood: The major longitudinal studies* (pp. 48–70). New York: Guilford.

*Stams, G. J., Juffer, F., & van IJzendoorn, M. H. (2002). Maternal sensitivity, infant attachment, and temperament in early childhood predict adjustment in middle childhood: The case of adopted children and their biologically unrelated parents. *Developmental psychology*, *38*, 806–821.

*Suess, G. J., Grossmann, K., & Sroufe, L. A. (1992). Effects of infant attachment to mother and father on quality of adaptation in preschool: From dyadic to individual organisation of self. *International Journal of Behavioral Development*, *15*, 43–65.

Suomi, S. J. (2003). Gene-environment interactions and the neurobiology of social conflict. In J. A. King, C. F. Ferris, & I. I. Lederhendler (Eds.), *Roots of mental illness in children* (pp. 132–139). New York: New York Academy of Sciences.

Sutton, A. J., Duval, S. J., Tweedie, R. L., Abrams, K. R., & Jones, D. R. (2000). Empirical assessment of effect of publication bias on meta-analyses. *British Medical Journal*, *320*, 1574–1577.

Tabachnik, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston: Allyn & Bacon.

*Turner, P. (1991). Relations between attachment, gender, and behavior with peers in the preschool. *Child Development*, *62*, 1475–1488.

van IJzendoorn, M. H. (1997). Attachment, emergent morality, and aggression: Toward a developmental socioemotional model of antisocial behaviour. *International Journal of Behavioral Development*, *21*, 703–727.

*van IJzendoorn, M. H., Sagi, A., & Lambermon, M. W. E. (1992). The multiple caretaker paradox: Data from Holland and Israel. In R. C. Pianta (Ed.), *Beyond the parent: The role of other adults in children's lives* (pp. 5–24). San Francisco: Jossey-Bass.

van IJzendoorn, M. H., Schuengel, C., & Bakermans-Kranenburg, M. J. (1999). Disorganized attachment in early childhood: Meta-analysis of precursors, concomitants, and sequelae. *Development and Psychopathology*, *11*, 225–249.

*Van Zeijl, J., Mesman, J., Van Ijzendoorn, M. H., Bakermans Kranenburg, M. J., Juffer, F., Stolk, M. N., et al. (2006). Attachment-based intervention for enhancing sensitive discipline in mothers of 1- to 3-year-old children at risk for externalizing behavior problems: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *74*, 994–1005.

Verschueren, K., & Marcoen, A. (1999). Representation of self and socioemotional competence in kindergartners: Differential and combined effects of attachment to mother and father. *Child Development*, *70*, 183–201.

Vierikko, E., Pulkkinen, L., Kaprio, J., Viken, R., & Rose, R. J. (2003). Sex differences in genetic and environmental effects on aggresion. *Aggressive Behavior*, *29*, 55–68.

*Vondra, J. I., Shaw, D. S., Swearingen, L., Cohen, M., & Owens, E. B. (2001). Attachment stability and emotional and behavioral regulation from infancy to preschool age. *Development and Psychopathology*, *13*, 13–33.

Waters, E., & Deane, K. E. (1985). Defining and assessing individual differences in attachment relationships: Q-methodology and the organization of behavior in infancy and early childhood. In I. Bretherton & E. Waters (Eds.), Growing points of attachment theory and research. *Monographs of the Society for Research in Child Development*, *50*(1–2, Serial No. 209), 41–65.

Weaver, I. C. G., Cervoni, N., Champagne, F. A., D'Alessio, A. C., Sharma, S., Seckl, J. R., et al. (2004). Epigenetic programming by maternal behavior. *Nature Neuroscience*, *7*, 847–854.

*Weiss, S. J., & Seed, M. S. (2002). Precursors of mental health problems for low birth weight children: The salience of family environment during the first year of life. *Child psychiatry and human development*, *33*, 3–27.

*Wood, J. J., Emmerson, N. A., & Cowan, P. A. (2004). Is early attachment security carried forward into relationships with preschool peers? *British Journal of Developmental Psychology*, *22*, 245–253.

Child Development, March/April 2010, Volume 81, Number 2, Pages 457–463

Not Doing What You Are Told: Early Perseverative Errors in Updating Mental Representations via Language

Patricia A. Ganea
Boston University

Paul L. Harris
Harvard University

This research examined the ability of young ($N = 96$) children to learn about a change in the location of a hidden object, either via an adult's verbal testimony or from direct observation. Thirty-month-olds searched with equal accuracy whether they were told about the change or directly observed it. By contrast, when 23-month-olds were told about the change of location, they often returned to the container where they had last observed the object—even when that container was visibly empty. When interference from prior observational encoding was minimized, 23-month-olds, and even 19-month-olds, successfully updated their knowledge of the object's location on the basis of language. The processing demands of updating experience-based representations from new verbal information are discussed.

A century of research has revealed marked developmental changes in children's ability to update their knowledge of an object's location. We review those developmental changes and then discuss an important gap in our understanding of such updating.

One-year-old infants can watch an object being hidden and search correctly some 5–10 s later (Hunter, 1917). Indeed, they search correctly at successive locations, provided the delay between hiding and search trial is brief (Diamond, 1985; Harris, 1973). In the 2nd year, infants can keep track not just of visible displacements but also of invisible displacements. If an object is carried in a container (Haake & Somerville, 1985; Piaget, 1954), infants search at potential hiding places along the route of the transported object. Toddlers can even decode symbolic information about an object's displacement. For example, if a miniature toy dog is hidden at a particular location in a scale model of a room, 36-month-olds—but not 30-month-olds—treat that information as a representation of where a larger toy dog can be found in a full-size room (DeLoache, 1987, 2004).

Thus, there are major developments in the information that children use to update their knowledge of an object's location: 12-month-olds process an object's visible displacements; 18-month-olds process hidden displacements signaled by movements of a container; 3-year-olds decode the implications of a symbolic displacement—the movement of an object within a scale model. These changes suggest an emerging capacity to entertain more than one mental model of the world. Thus, theorists have proposed that infants entertain a single model but during the 2nd and 3rd years start to process more than one model: They compare successive sightings of a container to infer the hiding place of an object or they map the location of an object within a scale model to its location in a larger space (Perner, 1991; Suddendorf & Whiten, 2001).

Despite these changes, we know little about children's ability to update their knowledge of an object's location via language. Yet children do update on the basis of verbal input well before they use the symbolic information in a scale model. For example, having learned the name for a toy, toddlers went to an adjacent room where they were told that it had undergone a change (become wet). On their return, 22-month-olds, but not 19-month-olds, indicated the wet toy as opposed to a dry exemplar when asked for it by name (Ganea, Shutts, Spelke, & DeLoache, 2007).

We thank the children who participated and their parents. We also thank Jasmine DeJesus, Carina Wind, and the undergraduate research assistants at the Language & Learning Lab at Boston University for their help with data collection and coding.

Correspondence concerning this article should be addressed to Patricia A. Ganea, Department of Psychology, 64 Cummington St., Boston University, Boston, MA 02215. Electronic mail may be sent to pganea@bu.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8102-0003

We modified this paradigm to compare children's ability to update their knowledge of an object's location on the basis of verbal testimony or direct observation. In the testimony condition, we asked if 23- and 30-month-olds who had seen an object placed in location A would search at location B, if they were told—during their brief absence from the room—that the object had been moved from A to B. In the direct observation condition, children observed the change of location for themselves.

We expected toddlers to successfully search for the object at the new location in the observation condition. This corresponds to the classic "A-not-B" search paradigm, which children solve by the middle of the 2nd year. In the testimony condition, by contrast, two possible outcomes were anticipated. Children might update their initial representation of the object's location on the basis of language and correctly search for it at B. Such competence would be consistent with the findings of Ganea et al. (2007).

Alternatively, children might have difficulty with verbal updating, particularly when the revision of a prior, observation-based representation is required (Zelazo, Sommerville, & Nichols, 1999). On this hypothesis, children should initially search at A despite having been told that the object is at B. Only when prompted to search further might children turn from A to B. The executive difficulties faced by infants aged 9–12 months in searching at an object's most recent location, particularly in the face of interference from a prior representation, have long been entertained as an explanation for errors in the A-not-B search paradigm (Aguiar & Baillargeon, 2000; Diamond, 1985; Harris, 1973; Marcovitch & Zelazo, 1999, 2009). We anticipated that such difficulties might re-emerge if information about an object's new location were provided via language rather than direct observation.

Study 1

Method

Participants

Twenty 23-month-olds (9 girls, range = 22.4–24.7, *M* = 23.5) and sixteen 30-month-olds (9 girls, range = 28.3–33.3, *M* = 30.6) were tested. Five children were excluded—for fussiness (4) or lack of English proficiency (1). The majority were White, middle class, and English-speaking.

Materials

Four hiding locations were used: one green pillow on top of a brown couch, a three-drawer cabinet, a blue ottoman (called "chair"), and a wicker basket. For 30-month-olds, the basket was replaced by a beige box with a removable lid, a green cloth bag was placed on top of the ottoman (referred to as "bag"), and the green pillow was replaced with a black pillow. These changes were made because the 30-month-olds were overly interested in the basket during pilot testing and seemed confused when the experimenter called the ottoman a chair.

An opaque curtain was hung in the room to create an inner space (218 × 113 in.) containing the hiding locations and an outer space (34 × 113 in.; see Figure 1). A transparent plastic window was inserted in the curtain (8.5 × 11 in.), 44 in. from the

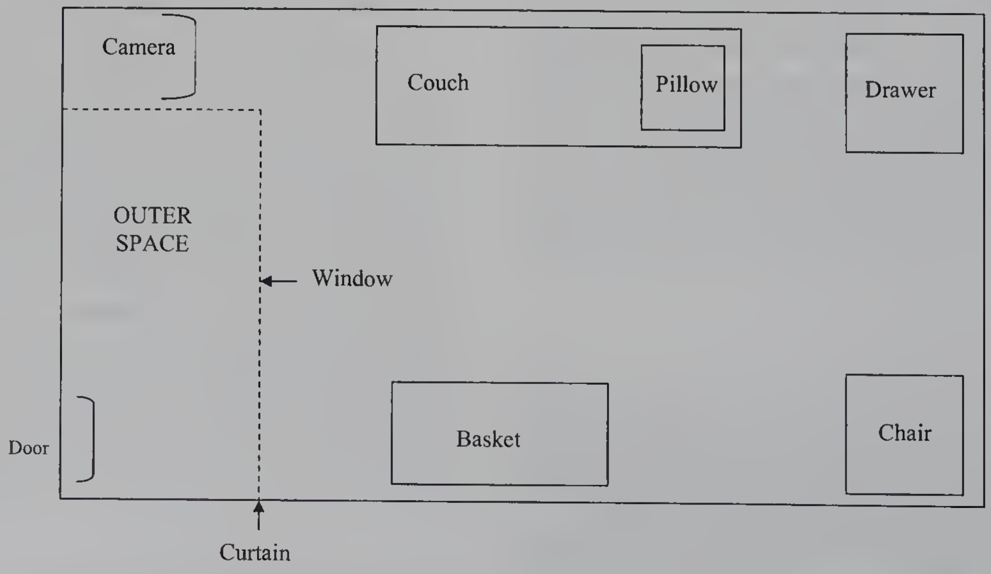


Figure 1. Illustration of experimental room layout.

ground. Children could not see through it unless lifted by an adult. Two stuffed animals served as hiding objects: a black and white puppy and a brown monkey. One camera was placed near the curtain to record the experimental session.

Procedure

The 23-month-olds received the two conditions (testimony and direct observation) in counterbalanced order. The 30-month-olds received the testimony condition first (during piloting children were distracted during the testimony if they had previously watched through the window in the direct observation condition). Each condition was presented following a familiarization phase.

Familiarization phase. Two experimenters sat with the child in the inner space, with the parent in the outer space. The familiarization phase was designed to ensure that children were aware of the four hiding locations and remembered the initial hiding location of the toy. To introduce the child to the hiding locations, the experimenters played a hiding game. One experimenter (E1) hid a toy with the child in one of the four locations (e.g., behind the pillow), while the other experimenter (E2) covered his or her eyes. After the hiding, E1 asked the child to show E2 the location of the toy. This game was repeated for all four locations.

E1 and the child then hid the toy a final time. E1 asked the child to point out each of the four hiding locations (e.g., ''Can you show me the basket?'') and the hidden animal (e.g., ''Can you show me where the puppy is?''). If the child did not indicate the animal's location, E1 showed the child where the animal was, returned it to its location, and asked the child to identify the location again. If the child failed a second time, the session continued but the child's data were excluded. Then E1, the child, and the parent moved behind the curtain for the test phase.

Test phase. The child participated in two conditions—testimony and direct observation. In both conditions, E2 moved the toy from its hidden location diagonally across the room to a new location (e.g., from the pillow to the chair). In the testimony condition, E1 looked through the window and told the child about the change in location: ''Look *E2* is moving the puppy! *E2* is moving the puppy from the pillow to the chair. The puppy is behind the chair. Yeah the puppy is behind the chair!'' In the direct observation condition, the child was lifted up to the window and told: ''Look *E2* is moving the puppy! *E2* is moving the puppy to a new place. Now the puppy is in a new place!'' Note that the verbal information in the direct observation condition did not identify the new location of the toy, only the fact that the toy was moved. For 30-month-olds, E1 used less verbal input during the direct observation condition—''Let's see what *E2* does in there. Do you see her? Look!'' The input given to younger children was judged to be redundant for older children. Then, E1 asked the child to get ready to find the toy and opened the curtain to allow the child to search. If the child did not find the toy on the first try and did not continue to search, she or he was prompted to continue searching (''Do you remember that she moved the puppy? Where is it?'').

The procedure was repeated for the other test condition with the other stuffed animal. As in the previous condition, E1 hid the toy and asked the child to point out the four hiding locations and the location of the hidden toy. After the child identified the toy's location, E1 proceeded with the test phase. The hiding locations for this phase were the remaining two not used in the first trial and the child received the opposite condition from the first trial. Across the three studies reported here, there was no effect of order of conditions on search behavior.

Coding

Films of test sessions for all three studies were coded to identify children's search behavior. Interrater agreement ranged from 94% to 100% (Cohen's kappa = .85–1.00). Disagreements were resolved by another person.

Results and Discussion

We used chi-square or binomial tests to compare children's search in the two conditions with chance levels (for the first response, chance was set at .25 because there were four possible locations to search; when children searched a second time after an error on their initial search, chance was .33).

When children directly observed the change in location, both younger (20 of 20) and older children (14 of 16) searched correctly (binomial tests, $p < .01$; see Figure 2). However, when children were told about the change, few of the younger (4 of 20) but most of the older (12 of 16) children used the verbal information to search correctly ($p < .01$). A chi-square test confirmed that the proportion of children searching correctly in the testimony condition was larger in the older than in the younger group, $\chi^2(1, 36) = 8.77, p < .001$.

Of the 16 younger children who failed to search correctly, most (87%) searched in the initial location

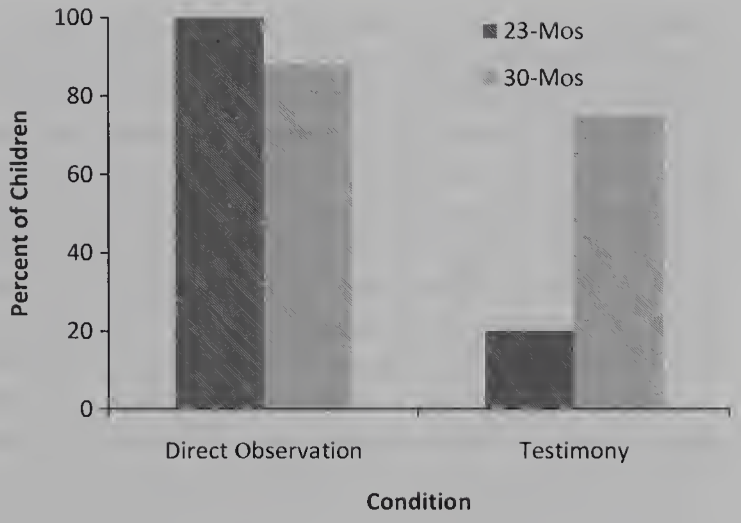


Figure 2. Percentage of children in each of the two conditions who searched correctly in the new location in Study 1.

of the toy ($p < .01$). On their second search attempt, 50% of children searched correctly. Thus, some children may have successfully encoded the new location. However, caution is needed in interpreting this finding because the number of correct second choices was not significantly different from chance, $\chi^2(1, 16) = 2.09$, $p = .14$.

Why did most 23-month-olds not use the verbal information for their initial search? According to the second hypothesis described earlier, toddlers' utilization of verbal information about the new location was prone to interference from the visual representation of the object's initial location. However, children may not have understood the testimony in the first place. This interpretation seems unlikely because Ganea et al. (2007) showed that 23-month-olds can understand verbal information about an invisible change. However, testimony about a location change may be harder to understand than testimony about a state change (i.e., dry to wet).

Study 2 was designed to discriminate between these two interpretations. At the end of the familiarization phase, the toy was left in the middle of the room. If children's utilization of verbal testimony in Study 1 was prone to interference from a prior visual representation of the object's specific location, interference should be reduced in Study 2 because there was no specific prior location. Thus, if children can understand verbal information about a toy's displacement, they should search correctly. However, if children cannot understand information about a location change (as opposed to a state change) they should fail to search correctly.

Because 30-month-olds responded accurately in both the visual and verbal conditions of Study 1, they were not included in Study 2. Instead, to probe the origins of toddlers' updating skills, 19-month-olds were also tested. Recall that the 19-month-olds tested by Ganea et al. (2007) did not respond systematically when given information about a state change.

Study 2

Method

Participants

Sixteen 23-month-olds (9 girls, range = 22.0–24.4, $M = 23.0$) and sixteen 19-month-olds (9 girls, range = 18.3–20.9, $M = 19.4$) were tested. Ten children were excluded—for fussiness (8) or lack of English proficiency (2).

Materials

Materials were the same as in Study 1, except that only two hiding locations were used for the 19-month-olds. The 23-month-olds were tested as in Study 1 with four hiding locations available in the room.

Procedure

Familiarization phase. Children were introduced to the testing locations and asked to identify them as in Study 1. At the end of this phase, E1 left the toy in the middle of the room.

Test phase. After familiarization, E1, the child, and the parent moved behind the curtain. Twenty-three-month-olds then received each of two test conditions—testimony and direct observation. Nineteen-month-olds were tested only in the testimony condition. In both conditions, E2 moved the toy from the middle of the room to a specific location. In the testimony condition, E2 came into the outer space and told the child about the new location: "Guess what? I moved the puppy! I put the puppy behind the pillow. Now puppy is behind the pillow." E1 then reinforced E2's testimony: "Did you hear that? *E2* put the puppy behind the pillow!" In the direct observation condition, E1 simply told the child to watch the change in location through the window in the curtain: "Let's see what *E2* does in there. Do you see her? Look!"

Results and Discussion

The majority of 23-month-olds searched correctly in both the testimony (12 of 16 children, $p < .01$)

and direct observation condition (14 of 16 children, $p < .01$). Thus, when the verbal information about the new location was not vulnerable to interference from visual information about the object's prior location, most 23-month-olds searched correctly. Indeed, most 19-month-olds (13 of 16, $p < .05$) also searched correctly based on testimony.

Study 2 supports two conclusions. First, toddlers can understand and use verbal testimony about an absent object's location. Second, when considered in relation to the perseverative errors observed in Study 1, the findings from Study 2 suggest that 23-month-olds have difficulty acting on testimony that is vulnerable to interference from specific previously encoded visual information. Study 3 was designed to probe this vulnerability.

In Study 3, a new group of 23-month-olds was tested. The initial location of the toy was either an open or a closed container. In the open container condition, children could see that the toy was not in the container upon re-entry.

Study 3

Method

Participants

Twenty-eight 23-month-olds (12 girls, range = 22.0–24.4, $M = 23.3$) were tested. Five children were excluded—because of fussiness (3), lack of English proficiency (1), and parental interference (1).

Materials

The hiding locations were the same as in Study 1, but the box was altered. In the closed box condition, an opaque box with a removable lid was used. In the open box condition, the box had a hole cut in its front panel so that its contents remained visible.

Procedure

With few exceptions, the procedure was the same as for Study 1. First, before the test phase, the initial hiding location of the toy was always in the box (either open or closed, depending on the condition). The location of the box was varied so that in each condition, the box was placed either in the front or in the back of the room, and either on the right or on the left side. Box location was counterbalanced both across trials and across children. Because the box had a different location on each trial and to avoid using the previous location of the box on the next trial, one hiding location was removed. (In this study there were only three hiding locations, bringing the chance level for the first search to .33.) The box was placed on a furniture item (drawer set, couch, ottoman, small table), so that it was at the child's eye level when entering the room.

Children received *testimony* that the toy was moved in both the open and closed conditions, counterbalanced for order. After E2 moved the toy, she told the child about the new location (e.g., "Guess what? I moved the puppy! I moved the puppy from the box to the bag. Now the puppy is in the bag"). E1 reinforced E2's testimony and asked the child to get ready to find the toy. If the child did not find the toy and did not continue to look for it, she or he was prompted to continue searching. The procedure was repeated with the other box (open or closed) and stuffed animal for a second trial.

Results and Discussion

Few children correctly searched in the new location in either the closed (2 of 28) or open box condition (8 of 28). A Fisher exact test confirmed that there was no significant difference in the rate of errors across conditions ($p = .09$). Children who failed to search correctly mostly searched in the old location of the toy: 92% in the closed box condition and 75% in the open box condition, $p < .05$ (chance level .33). Children who failed to update (20) in the open box condition exhibited four types of search: the majority (15) went to the open box and looked inside or searched in it, some (3) approached the open box but halfway switched direction to another location, one child looked at the box upon entry but did not approach it, and finally, one child went straight to an incorrect location without looking at the old location. Of the children who failed to search in the closed box condition (26), the majority (24) went all the way to the box and searched in it, whereas 2 children went to another location in the room.

On their second attempt to find the toy, over half of the children searched correctly (compared to a chance level of .50): 18 of 26 (69%) in the closed box condition, $p < .05$; 14 of 19 (74%) in the open box condition (one child did not search a second time), $p < .05$, indicating that children retained verbal input regarding the new location of the toy despite not searching there immediately. Figure 3 illustrates 23-month-olds' performance across the three studies.

To summarize, 23-month-olds have difficulty in updating their representation of an object's location despite visible evidence that their prior visual encoding is no longer valid. In the open box condition, the majority of children approached the visibly empty container and many searched in it.

General Discussion

Study 1 showed that 30-month-olds were able to search accurately whether they observed the object's change of location or were told about it. Twenty-three-month-olds, by contrast, searched correctly only after observing the change of location. When told about it, they often searched where they last observed the object. Study 2 showed that children could take in verbal information about a new location. When the object was left in the middle of the room, both 19- and 23-month-olds searched correctly when told that it had been moved. Study 3 underlined the lure of earlier observation-based encoding. Twenty-three-month-olds made perseverative errors not only when the initial hiding place was a closed box (replicating the results of Study 1) but also when it was an open box. They could see that the open box was empty yet still approached it, ignoring what they were told about the object's new location. Indeed, many children searched inside the visibly empty box.

The pattern of error displayed by 23-month-olds is reminiscent of the pattern shown by younger infants in the A-not-B paradigm in two notable respects. In that paradigm, infants also return to B even when it is visibly empty (Harris, 1974). Second, when given an opportunity to search after an error, infants display evidence of having encoded the new location. They search there rather than elsewhere (Webb, Massar, & Nadolny, 1972). By implication, children's difficulties in updating on the basis of language are part of a more general difficulty in updating. Nevertheless, the 23-month-old children in this study could update on the basis of visual information. How does language-based updating impose different cognitive demands compared to visually based updating?

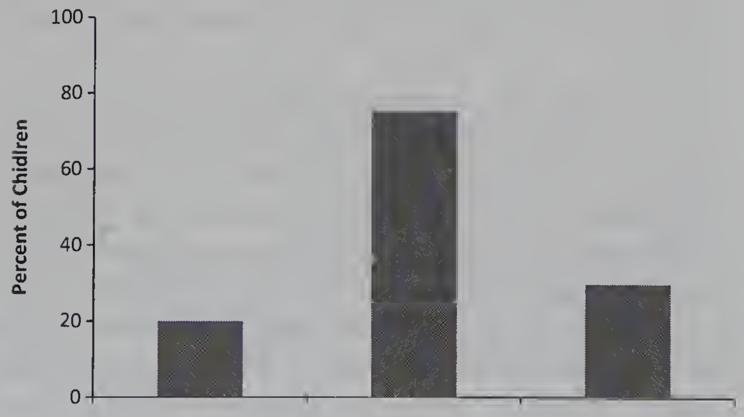


Figure 3. Percent of 23-month-old children in each study who searched in the new location in the testimony condition.

Recent studies with adults suggest that the cognitive mechanisms involved in the perception of action are also involved in processing action descriptions (Yaxley & Zwaan, 2007; Zwaan & Taylor, 2006). Thus, spatial imagery is automatically activated during language processing (Barsalou, Simmons, Barbey, & Wilson, 2003; Bergen, Linday, Matlock, & Narayanan, 2007; Fischer & Zwaan, 2008; Spivey & Geng, 2001; Zwaan, Stanfield, & Yaxley, 2002). Currently, there is little information about the emergence of children's ability to activate such imagery when processing verbal descriptions. Thus, we do not know whether the children tested in our study imagined the transformation of the toy's location when the updating involved language. What our results do indicate is that at 30 months—but not at 23 months—comprehension of action sentences ("The toy was moved from A to B") is equivalent to action observation (seeing the toy moved from A to B).

What changes take place in language-based updating between 23 and 30 months?

First, language-based representation of the new location might increase in strength relative to the visually based representation of the initial location. By analogy, in the classic A-not-B paradigm, infants initiate an action toward the hidden object only when they have a strong representation of the object (Munakata, 2001; Munakata, McClelland, Johnson, & Siegler, 1997). In the current studies, when the basis for updating was visual information, children could form a strong representation of the new location and disregard the initial location. However, when the basis for updating was verbal information, younger children were lured to location A, even when it was visibly empty, suggesting that their visual representation of that location was stronger than the language-based representation of the new location.

A second explanation is that updating in the testimony condition required integration of information from two modalities—visual and verbal. In the direct observation condition, the updating required integration only within the visual modality. In the testimony condition, children had to revise a visual representation on the basis of verbal information. Arguably, younger children's ability to integrate information across different modalities is less efficient than that of older children. One way to test this hypothesis is to manipulate both input (verbal or visual) and initial representation (verbal or visual) for both age groups.

Finally, children's ability to *edit* a prior visually based representation on the basis of subsequent verbal information may change between 23 and 30 months. The editing process may be straightforward when it involves only the addition of new information, as in Study 2 (''The object is now in B'') or as in Ganea et al. (2007; ''Lucy is now wet''), but more demanding if it involves deletion and replacement of specific information as in Studies 1 and 3 (''The object was in location A. Now it is in location B''). In Study 2, children may have stored a generic representation that the toy was ''in the room'' so that subsequently, they only needed to supplement that generic representation. In Studies 1 and 3, the toy was placed in a particular location so that updating called for deletion of that specific initial location and replacement with information about its new location. Developmental differences might be expected, depending on whether updating involves simple addition or deletion and replacement.

In conclusion, toddlers can update on the basis of verbal testimony by 19 months—almost at the onset of language production. Nevertheless, even at 23 months the updating process is error prone. Echoing the difficulties shown by preverbal infants in shifting from one hiding place to another, toddlers display perseverative errors when they have been told about a shift in location instead of observing it for themselves.

References

Aguiar, A., & Baillargeon, R. (2000). Perseveration and problem solving in infancy. In H. W. Reese (Ed.), *Advances in child development and behavior* (Vol. 27, pp. 135–180). San Diego, CA: Academic Press.

Barsalou, L. W., Simmons, W. K., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, *7*, 84–91.

Bergen, B. K., Lindsay, S., Matlock, T., & Narayanan, S. (2007). Spatial and linguistic aspects of visual imagery in sentence comprehension. *Cognitive Science*, *31*, 733–764.

DeLoache, J. S. (1987). Rapid change in the symbolic functioning of very young children. *Science*, *238*, 1556–1557.

DeLoache, J. S. (2004). Becoming symbol-minded. *Trends in Cognitive Science*, *8*, 66–70.

Diamond, A. (1985). Development of the ability to use recall to guide action, as indicated by the infant's performance on AB. *Child Development*, *56*, 868–883.

Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: A review of the role of the motor system in language comprehension. *Quarterly Journal of Experimental Psychology*, *61*, 825–850.

Ganea, P. A., Shutts, K., Spelke, E., & DeLoache, J. S. (2007). Thinking of things unseen: Infants' use of language to update mental representations. *Psychological Science*, *18*, 734–739.

Haake, R. J., & Somerville, S. C. (1985). Development of logical search skills in infancy. *Developmental Psychology*, *21*, 176–186.

Harris, P. L. (1973). Perseverative errors in search by young infants. *Child Development*, *44*, 28–33.

Harris, P. L. (1974). Perseverative search at a visibly empty place by young infants. *Journal of Experimental Child Psychology*, *18*, 535–542.

Hunter, W. S. (1917). The delayed reaction in a child. *Psychological Review*, *24*, 74–87.

Marcovitch, S., & Zelazo, P. D. (1999). The A-not-B error: Results from a logistic meta-analysis. *Child Development*, *70*, 1297–1313.

Marcovitch, S., & Zelazo, P. D. (2009). A hierarchical competing systems model of the emergence and early development of executive function. *Developmental Science*, *12*, 1–25.

Munakata, Y. (2001). Graded representations in behavioral dissociations. *Trends in Cognitive Sciences*, *5*, 309–315.

Munakata, Y., McClelland, J. L., Johnson, M. H., & Siegler, R. S. (1997). Rethinking infant knowledge: Toward an adaptive process account of success and failures in object permanence tasks. *Psychological Review*, *104*, 686–713.

Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.

Piaget, J. (1954). *The child's construction of reality*. New York: Basic Books.

Spivey, M. J., & Geng, J. J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research*, *65*, 235–241.

Suddendorf, T., & Whiten, A. (2001). Mental evolution and development: Evidence for secondary representation in children, great apes, and other animals. *Psychological Bulletin*, *127*, 629–650.

Webb, R. A., Massar, B. M., & Nadolny, T. (1972). Information and strategy in the young child's search for hidden objects. *Child Development*, *43*, 91–104.

Yaxley, R. H., & Zwaan, R. A. (2007). Simulating visibility during language comprehension. *Cognition*, *105*, 229–236.

Zelazo, P. D., Sommerville, J. A., & Nichols, S. (1999). Age-related changes in children's use of external representations. *Developmental Psychology*, *4*, 1059–1071.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, *13*, 168–171.

Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General*, *135*, 1–11.

Child Development, March/April 2010, Volume 81, Number 2, Pages 464–471

"Who Said You Could Wear My Sweater?" Adolescent Siblings' Conflicts and Associations With Relationship Quality

Nicole Campione-Barr
University of Missouri–Columbia

Judith G. Smetana
University of Rochester

A new measure of sibling conflict was used to identify 2 types of conflicts in 115 adolescent sibling pairs (older siblings, $M = 15.59$, $SD = 2.01$ years; younger siblings, $M = 13.02$, $SD = 2.06$ years). Conflicts overall were more frequent than intense and more likely to involve the invasion of the personal domain than conflicts involving equality and fairness, especially by early adolescent older siblings. For both siblings, and with parents' ratings of their relationship with the target child controlled, these conflicts were negatively associated with sibling relationship quality. The implications of the findings for sibling relationships during adolescence are discussed.

Siblings have frequent disagreements during adolescence. Furman and Buhrmester (1985) found that of six major relationships (mothers, fathers, siblings, grandparents, friends, and teachers), early adolescents perceived conflict as occurring most frequently with siblings, perhaps due to the nature of the relationship. Like other family members, relationships with siblings are based on kinship and long histories of interactions; thus, the relationship can withstand conflict without fear of dissolution. However, adolescent sibling conflicts have garnered relatively little research attention.

A meta-analysis has demonstrated that parent–adolescent conflict frequency peaks during early adolescence (Laursen, Coy, & Collins, 1998). Mirroring these results, Buhrmester and Furman (1990) found that among 3rd, 6th, 9th, and 12th graders, conflict with siblings was greatest at 6th grade. This could be due to increases in autonomy seeking, especially as increased sibling conflict was reported mostly by early adolescent younger siblings (Buhrmester & Furman, 1990). Younger siblings may see older siblings as similar to parents in that there is a hierarchy in the relationship. As siblings grow older and developmental differences between them decline, older siblings may hold on to earlier patterns of asymmetry in the relationship, which may be bothersome to younger siblings. Therefore, early adolescents may feel as though some autonomy must be gained from siblings as well as parents to emphasize their maturity.

Buhrmester and Furman (1990) also found that older children reported that their sibling relationships were more egalitarian than did younger children, with adolescents reporting less dominance by their siblings than children. Younger siblings may become more autonomous and no longer allow older siblings to dominate the relationship. However, companionship and affection were also lower in the older age groups. Buhrmester and Furman concluded that sibling relationships become less intense, both positively and negatively, as children move into adolescence. Finally, these researchers found that same-sex sibling dyads felt closer than opposite-sex dyads.

Investigations of specific processes involved in sibling relationships across all age groups, as well as adolescent siblings more generally, have been limited until recently (e.g., Feinberg, McHale, Crouter, & Cumsille, 2003; Kim, McHale, Osgood, & Crouter, 2006). Research on sibling relationships

---

The research described here was conducted in partial fulfillment of the requirements of the Ph.D. degree for the first author at the University of Rochester. We would like to thank Dr. Kenneth Hilton of the Rush-Henrietta School District and the many families who participated in this research. We would also like to thank Denise Gettman, Aaron Metzger, Marina Tasopoulos-Chan, and Myriam Villalobos, as well as our undergraduate research assistants for their assistance with family visits and data entry. We thank the Fetzer Institute for their support of this research.

Correspondence concerning this article should be addressed to Nicole Campione-Barr, Department of Psychological Sciences, University of Missouri–Columbia, 210 McAlester Hall, Columbia, MO 65211. Electronic mail may be sent to campionebarrn@missouri.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8102-0004

in early and middle childhood has examined some aspects of sibling conflicts such as resolution strategies (e.g., Howe, Rinaldi, Jennings, & Petrakos, 2002), parental intervention (Ross, Filyer, Lollis, Perlman, & Martin, 1994), and, to a much lesser extent, the content of sibling conflicts (McGuire, Manke, Eftekhari, & Dunn, 2000). McGuire et al. (2000) found that for sibling pairs in middle childhood, the most common issues of conflict involved sharing of personal possessions, and these issues were more common than conflicts over household items. The issues that cause disagreement between adolescent sibling dyads have not been studied, however, and this was one purpose of the present study.

Research from the social domain perspective (see Smetana, 2006) has investigated differences in conflict between parents and adolescents as a function of the topic of the conflict (Smetana & Asquith, 1994). This research shows that parents and adolescents agree that parents should have authority to regulate *moral* (pertaining to others' rights, welfare, or fairness) and *conventional* (contextually relative, arbitrary social norms) issues throughout adolescence, whereas children should have some authority over *personal* issues (pertaining to privacy, control over their bodies, and preferences and choices) well before adolescence (Smetana, 1989; Smetana & Asquith, 1994). However, conflicts are typically over issues that cross the boundaries of different social domains (i.e., issues that adolescents view as infringing on their personal jurisdiction that parents view as conventional or prudential).

The importance of the personal domain during adolescence and McGuire et al.'s (2000) findings for younger sibling pairs suggest that conflicts between siblings also may pertain to personal issues and adolescents' perceptions that their sibling has infringed on their personal domain (e.g., one's personal space or friends). Given the age-related increases in appeals to personal jurisdiction in adolescents' conflicts with parents (Smetana, 1989), sibling conflicts over personal issues may be more frequent and more heated than moral conflicts involving the equality and fairness of household resource distribution (e.g., computer use), which also may occur between adolescent siblings.

The present study examined specific types of conflict among adolescent siblings using a newly developed measure that is similar in structure and response choices to the Issues Checklist (Robin & Foster, 1989), a frequently used measure of adolescent–parent conflict. Previous research has shown that it is important to consider both conflict frequency and its intensity given that they differentially impact relationships and development (Montemayor, 1983). Therefore, both forms of conflict were examined in the present study, and it was hypothesized that issues involving invasion of the personal domain would be more frequent and more intense than those involving moral issues of equality and fairness. Additionally, we investigated the association between the different types of conflicts and sibling relationship quality. Based on previous research showing that adolescents perceived parents to be overly controlling when they tried to regulate adolescents' personal domain (Smetana & Daddis, 2002), we hypothesized that conflicts involving threats to one's sense of self or personal relationships (e.g., copying what the other does, hanging around when the other has friends over) would be associated with lower levels of trust and communication with siblings, whereas conflicts involving equal distribution of resources (e.g., distribution of chores) would be less heated and have less influence on the relationship than conflicts about personal issues. Also, parents' reports of relationship quality with their children were included in these analyses to ensure that significant effects were due to the sibling relationship and not general levels of closeness within the family.

Method

Participants and Sample

Participants were 115 sibling pairs and one of their parents (108 mothers and 7 fathers). The sibling pairs were within one to four grade levels of one another. Older siblings in the dyads ranged in age from 12.03 to 20.38 years ($M = 15.59$, $SD = 1.98$); younger siblings ranged from 8.48 to 16.57 years ($M = 13.07$, $SD = 2.01$). The mean age difference was 2.53 years ($SD = .95$). Initially, 7th- and 10th-grade adolescents were recruited as part of a larger study; half of the students at each grade participated with an older sibling and half participated with a younger sibling. For the present study, sibling pairs were divided into three age cohorts. The preadolescent-early adolescent cohort ($n = 24$) consisted of 7th graders ($M = 10.23$ years, $SD = 1.02$) and their younger siblings ($M = 12.79$ years, $SD = .51$); the early-middle adolescent cohort ($n = 62$) consisted of early ($M = 12.93$, $SD = .75$) and middle ($M = 15.50$, $SD = 0.83$; some of whom were 7th graders with an older sibling, and some of whom were 10th graders with a younger sibling) adolescents. The middle-late adolescent cohort ($n = 29$) consisted of 10th graders ($M = 15.70$, $SD = .450$) and their older siblings ($M = 18.11$, $SD = 0.93$). Due to the small and uneven numbers of sister–sister and brother–brother sibling dyads when examined within the three age groups, we examined sibling sex composition as either same- or mixed-sex dyads ($n_{\text{same-sex}}$: pre-early = 8; early-middle = 35; middle-late = 17, $n_{\text{mixed-sex}}$: pre-early = 16; early-middle = 28; middle-late = 12). Age cohorts did not differ significantly by sex composition, mother's education, or family income; however, middle-late adolescent pairs were more likely than early-middle adolescent pairs to reside in single- or stepparent households.

The sample was primarily European American (88%), with the remaining participants African American (9%) or other ethnicities (3%). Most siblings lived with both biological parents (69%); the remainder resided in stepparent (9%) or single-parent families (21%). Family structure (examined as two-biological parent families vs. all else) and ethnicity (European American vs. all else) did not differ significantly by age cohort. The median yearly family income was between $55,000 and $69,999, and most parents (86%) had some college education. Mothers' and fathers' mean ages were 42.47 and 43.92 years (SDs = 4.69 and 5.65), respectively.

Measures

Sibling conflict. The Sibling Issues Checklist was developed for the present study, based on the format of the Issues Checklist (Robin & Foster, 1989); however, the content of the individual items was selected to be appropriate to siblings. The 20 issues were chosen based on a content analysis with a pilot sample of sixty-five 6th to 9th graders who were demographically similar to the study sample. They gave free responses to the question, "What types of things do you fight about with your brothers or sisters most often?" Responses suggested by more than 20% of the pilot sample were used in the final scale. As found in previous research (Smetana, 1989; Smetana & Asquith, 1994), the conflict topics reported by the pilot sample were mundane, everyday issues. Both older and younger siblings rated how often the issues were discussed (frequency) on a 5-point scale from 1 (*never*) to 5 (*very often*), and how "hot" the discussions got (intensity) on a 5-point scale ranging from 1 (*calm*) to 5 (*very angry*).

Relationship quality. The Trust and Communication subscales of the Armsden and Greenberg (1987) Parent and Peer Attachment Inventory were used to assess relationship quality. These scales have been used previously to assess relationships with mothers, fathers, friends, and romantic partners, but we adapted the questions to assess sibling relationship quality (e.g., My brother/sister accepts me as I am; My brother/sister can tell when I'm upset). As found previously, the Trust and Communication subscales were highly correlated, rs = .69 and .78 for older and younger siblings, respectively. Therefore, the two subscales were combined into a single mean relationship quality score. The scale consisted of 18 items rated on a 5-point scale ranging from 1 (*almost never or never*) to 5 (*almost always or always true*). Cronbach's alphas were .87 for both older and younger siblings' reports. Both siblings reported on their relationships with their participating sibling; sibling reports were moderately correlated, $r(115) = .52$, $p < .001$, suggesting that sibling dyads share some commonalties (but not complete overlap) in their views of their relationship. Finally, parents reported on their relationship with each of their two children using the original version of the inventory (αs = .80 for both).

Procedures

Participants were recruited from a Northeastern U.S. suburban school district. Letters describing the study were sent to parents of 7th- and 10th-grade students, and parents mailed back interest response sheets. Researchers also made presentations to students during their "homebase" classes to encourage their participation. Inclusion criteria were that students have a nontwin sibling less than 5 years older or younger than they were. Older siblings had to be first-born children, and younger siblings could be any later-born status, although most (80%) were second-born. As part of the larger study, families were paid honoraria for participating in a 2-hr visit at the university lab, during which they completed consent and assent forms, questionnaire measures, interviews, and sorting and family interaction tasks.

Results

Preliminary Analyses: Sibling Conflict Factors

The sibling conflict scales were empirically derived. To eliminate dependencies in the data and to determine whether the results replicated across siblings, we analyzed older and younger siblings' reports separately for ratings of frequency and intensity. This resulted in four separate principal components analyses with varimax rotation. Three inclusion criteria (eigenvalues greater than 1.00, scree tests, and meaningfulness and interpretability) were used in determining the number of factors. Variables that loaded .40 or greater on a factor, had low cross-loadings, and loaded highly on the same factor for at least three of the four analyses were retained in the scale.

The two hypothesized factors, Equality and Fairness (7 items: control over TV remote, TV show to watch, computer use, turn in front seat of car, which radio or CD to listen to, phone use, places to go with family), and Invasion of the Personal Domain emerged. Four of the six items loading on this latter scale entailed infringing on one's personal domain (going into other's room, copying what other does, borrowing without permission, and hanging around other's friends), but ''teasing and being mean'' (which pertains to a moral issue) and ''breaking other's things'' unexpectedly loaded on this scale. Seven of the 20 items (whose turn for chores, turn to play video games, rules of playing game, bathroom use, eating other's food, parents' attention, telling other what to do) were dropped from the scales due to cross-loading or inconsistent loading across the two scales (interested readers can write to the authors for the factor loadings and eigenvalues). Mean composites were created separately for younger and older siblings' reports of conflict frequency and intensity for the two types of issues. Alpha coefficients were acceptable for all scales (.77–.81).

Differences in Sibling Conflict by Form, Type of Issue, Dyad Age, and Sex Composition

Although ratings of conflict frequency and intensity were highly correlated, they were examined separately here, as they have been found to have different consequences for adolescent–parent relationships (Montemayor, 1983). A 2 (conflict rating: frequency vs. intensity) × 2 (reporter: older vs. younger sibling) × 2 (type of issue: fairness and equality vs. invasion of personal) × 3 (age cohort) × 2 (sex composition: same vs. mixed) mixed model analysis of variance (ANOVA), with conflict rating and type of issue as within-subjects factors and reporter, age, and sex as between subjects factors, was performed (see Table 1 for means).

Significant main effects of conflict rating, $F(1, 102) = 36.70$, $p < .001$, $\eta^2 = .27$, and conflict type, $F(1, 102) = 11.31$, $p < .001$, $\eta^2 = .10$, were qualified by a significant Rating × Type interaction, $F(1, 102) = 55.23$, $p < .001$, $\eta^2 = .35$; Rating × Reporter interaction, $F(1, 102) = 4.53$, $p < .05$, $\eta^2 = .04$; and a Rating × Type × Reporter interaction, $F(1, 102) = 10.12$, $p = .01$, $\eta^2 = .09$. Overall, conflict frequency ratings were higher than conflict intensity ratings,

Table 1
Means (Standard Deviations) of Conflict Frequency and Intensity by Reporter, Type, Age, and Sibling Gender

	Pre-early pairs		Early-middle pairs		Middle-late pairs	
	Same	Mixed	Same	Mixed	Same	Mixed
Frequency						
Personal						
Older	2.83 (1.20)	2.91 (1.30)	2.58 (0.93)	2.69 (0.80)	2.58 (0.87)	2.55 (0.94)
Younger	2.25 (1.06)	2.15 (1.07)	2.40 (0.84)	2.54 (0.88)	2.22 (0.91)	2.21 (1.06)
Equality						
Older	2.50 (0.83)	2.23 (0.89)	2.86 (0.98)	2.58 (0.77)	2.46 (0.99)	2.57 (0.85)
Younger	2.80 (1.27)	2.36 (0.70)	2.67 (0.86)	2.53 (0.95)	2.21 (1.01)	2.24 (0.86)
Intensity						
Personal						
Older	2.71 (0.98)	2.70 (1.27)	2.43 (0.85)	2.49 (0.84)	1.94 (0.94)	2.27 (1.04)
Younger	2.48 (1.31)	2.38 (0.96)	2.32 (0.91)	2.75 (0.97)	2.12 (0.92)	2.45 (1.20)
Equality						
Older	1.91 (0.67)	1.83 (0.86)	2.18 (0.77)	1.99 (0.72)	2.02 (0.87)	2.18 (0.75)
Younger	2.34 (1.11)	1.89 (0.69)	1.78 (0.66)	2.02 (0.88)	1.97 (1.07)	1.64 (0.74)

Note. Same = same sex; Mixed = mixed sex.

and conflicts were more likely to be over invasion of the personal domain than over equality and fairness. However, conflicts over equality and fairness were more frequent than intense, while there was no difference in rating of conflicts over invasion of the personal domain. Also, older siblings reported greater frequency of conflicts than younger siblings (there was no difference in intensity), and older siblings reported more frequent conflicts over invasion of the personal domain than younger siblings.

Finally, there were two significant three-way interactions with age cohort: Type × Reporter × Age Cohort, $F(2, 102) = 4.30$, $p < .05$, $\eta^2 = .08$, and Rating × Type × Age Cohort interaction, $F(2, 102) = 7.33$, $p < .001$, $\eta^2 = .13$. Older siblings reported more conflicts over invasions of the personal domain than over equality and fairness, but this was significant only in the youngest age cohort of siblings. Also, the finding that equality and fairness conflicts were more frequent than intense was only the case for the younger two age cohorts.

Associations of Type of Issue With Sibling Relationship Quality

To examine the association between conflict types and sibling relationship quality, four separate hierarchical regression analyses were performed on conflict frequency and intensity as rated by older and younger siblings. In the first step of the analyses, age cohort, sibling sex composition, and parental marital status were entered as controls (no other demographic variables were associated with the variables of interest; see Table 2). Additionally, in the first step, parents' report of their relationship with the older or younger child (depending on the analysis) was entered. Given that reports of relationship quality across multiple individuals within the same family are often moderately correlated, we controlled for parent–teen relationship quality so that any significant associations with sibling conflict were specifically associated with the sibling relationship, not the more general family environment. In the second step, the two types of conflict were entered simultaneously.

As shown in Table 3, for both conflict frequency and intensity, older siblings in same-sex pairs reported more trust and communication than those in mixed-sex pairs. Additionally, parents' trust and communication with younger siblings were significantly associated with older siblings' reports of trust and communication with these siblings. Finally, over and above the significant sex composition and parental relationship effects, conflicts over personal domain invasions, but not conflicts over equality and fairness, were negatively associated

Table 2
Correlations Among Demographic and Study Variables

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Sex															
2. Age	−.16														
3. Intact	.01	.07													
4. Educate	−.04	.00	−.10												
5. OFairFreq	−.14	.05	−.11	.14											
6. OPerFreq	.03	−.12	−.08	−.07	.58**										
7. OFairInt	−.07	.07	.02	.09	.68**	.48**									
8. OPerInt	.09	−.23*	.10	.14	.49**	.69**	.65**								
9. YFairFreq	−.08	−.08	−.24**	−.07	.38**	.33**	.40**	.28**							
10.YPerFreq	−.01	.02	−.19*	−.06	.22*	.34**	.23*	.21*	.53**						
11.YFairInt	.02	−.11	−.05	.01	.29**	.18	.39**	.21*	.57**	.39**					
12. YPerInt	.12†	−.07	−.06	−.04	.17	.26**	.19*	.20*	.37**	.77**	.49**				
13. ORelate	−.25**	−.02	.04	−.13	−.17	−.30**	−.24*	−.27**	−.07	−.20*	−.22*	−.27*			
14. YRelate	−.07	−.05	−.05	−.04	−.23*	−.18	−.26**	−.15	−.10	−.29**	−.23*	−.30**	−.52**		
15. PRelateO	−.01	−.19*	−.23*	−.21*	.06	−.02	−.02	−.04	.03	−.05	−.15	−.12	.25**	.32**	
16. PRelateY	−.02	−.24**	−.19†	−.15	.08	.01	−.03	−.02	.04	−.02	−.06	−.13	.28**	.22**	.71**

Note. Sex = sibling gender constellation (1 = same, 2 = mixed), Age = age cohort, Intact = 2 married parents versus all else, Educate = mother's education, O = older, Y = younger, Fair = equality and fairness, Per = personal domain, Freq = frequency, Int = Intensity.
$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$.

Table 3
Regression Analyses of Sibling Relationship Quality

	Older (frequency)			Older (intensity)			Younger (frequency)			Younger (intensity)		
	$F\Delta$	$R^2\Delta$	β	$F\Delta$	$R^2\Delta$	β	$F\Delta$	$R^2\Delta$	β	$F\Delta$	$R^2\Delta$	β
Step 1	4.84**	.15**		4.87**	.15**		3.61**	.12**		3.33*	.11*	
Age cohort			–.03			–.06			–.01			–.04
Sex comp.			–.26**			–.26**			–.08			–.07
Intact			.10			.13			–.05			.01
P relate			.30**			.28**			.31**			.27**
Step 2	6.82**	.10**		5.83**	.08**		4.88**	.08**		4.42*	.07*	
Equality			.06			–.11			.01			–.10
Personal			–.27*			–.21†			–.29**			–.21*

Note. P relate = parent relationship quality with the child being reported on; Intact = 2 married parents vs. all else.
$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$.

with older siblings' reports of relationships with younger siblings (significantly for frequency and approaching significance for intensity). The results of the analyses for younger sibling were similar. Whereas sex composition was not significant in the analysis, parents' reports of better relationships with older siblings were significantly associated with younger siblings' reports of better relationships with older siblings. Also, conflicts over invasion of the personal domain, but not conflicts over equality and fairness, were significantly and negatively associated with younger siblings' reports of relationship quality.

Discussion

The current study demonstrated that sibling conflicts in primarily White, middle-class adolescent dyads could be differentiated into personal domain conflicts or moral conflicts over equality and fairness of shared resources. Interestingly, conflicts over invasion of the personal domain also included a clearly moral issue—teasing and being mean—and a more pragmatic issue of property damage, suggesting that autonomy and space invasions shaded into (moral) concerns with welfare. Previous research by Helwig (2006) has demonstrated that especially in middle childhood and early adolescence, rights are often justified in terms of personal autonomy. Therefore, while conflicts over breaking one's personal possessions or teasing and being mean may pertain to moral rights, the adolescents may interpret these actions as invading their personal space or sense of agency. Nevertheless, conflicts involving the invasion of the personal domain (including psychological harm) were distinct in the factor analysis from moral conflicts over equality and fairness, and they were also more frequent and intense. These differences were most pronounced in early adolescent older siblings. This is somewhat consistent with previous findings by Buhrmester and Furman (1990) that sibling conflict is greater among early adolescents than younger children or older teens. The results also mirror the findings of Laursen et al.'s (1998) meta-analysis demonstrating that parent–adolescent conflict frequency peaks during early adolescence. Given that conflicts in both sibling and parent–adolescent relationships are mostly over personal issues, it is likely that both reflect adolescents' basic needs to claim some issues as outside of legitimate control by others. Claims to personal jurisdiction have been described as essential in establishing a sense of individuality (Nucci & Lee, 1993).

Buhrmester and Furman (1990) found that increased sibling conflict in early adolescence was mostly reported by younger siblings, whereas in the present study, early adolescent *older* siblings reported more conflicts, specifically over personal issues. These findings are consistent with McGuire et al.'s (2000) finding that middle-childhood older siblings reported more conflict over sharing personal items than did younger siblings. Together, these findings indicate the importance of assessing not only the frequency but also the content of sibling disputes. In the present study, early-adolescent older siblings reported having more conflicts over things such as borrowing something of the other's without asking, hanging around when the other has friends over, or copying what the other does than conflicts over issues such as bathroom use or whose turn it is to do chores. These types of conflicts may be more salient to older siblings in early adolescence, who may not want their younger siblings ''tagging along'' while they are trying to engage in more adult-like behaviors. An alternative explanation may be that as early adolescent older siblings begin the process of individuating from their family, they may engage in more conflicts with all family members, and this may be a change for younger siblings and parents, especially if they had previously enjoyed a close relationship.

We also found that for both older and younger siblings, more frequent and intense conflicts over invasion of the personal domain were associated with poorer relationships between siblings. Given that the data are cross-sectional, it is not clear whether this is because more conflicts over personal issues make it difficult to be in a close, trusting relationship, or whether siblings who have poor quality relationships feel free to more frequently invade one another's personal domain. Future research should include longitudinal samples to better understand the causal direction of this association.

Due to the way the sample was recruited, the age cohorts were uneven in size. Additionally, while the different gender constellations were relatively even, the cell sizes were too small to decompose same-sex dyads into male and female pairs or mixed-sex dyads into brother–sister and sister–brother pairs. Therefore, we only examined same sex versus mixed sex. Significant gender effects were found for relationship quality between siblings. Among older siblings, same-sex pairs reported being closer than mixed-sex pairs (there was no effect for younger sibling reports). Exploratory analyses suggested that, consistent with previous research, sister–sister dyads appeared to be closer than brother–brother dyads (*Ms* = 3.45, 3.20).

The present study was an important first step in examining the different types of issues that cause conflict between adolescent siblings and their associations with relationship quality. Although our measure was based on extensive free-response pilot testing, it is possible that we did not tap all of the salient issues in adolescent sibling conflicts. Thus, the measure needs to be validated through further research, particularly among ethnic minority and lower socioeconomic status youth, as well as youth from different cultures. Future research also should include observations of actual conflicts and examine how interactions, negotiation, and conflict resolution differ by the type of issue being discussed. Conflicts can be constructive and provide a context to learn negotiation or conflict resolution skills, but they also may be interpersonally destructive, and further research is needed to determine if this difference is dependent on the content of the conflicts or other aspects.

References

Armsden, G. C., & Greenberg, M. T. (1987). The Inventory of Parent and Peer Attachment: Individual differences and their relationship to psychological well-being in adolescence. *Journal of Youth and Adolescence, 16*, 427–454.

Buhrmester, D., & Furman, W. (1990). Perceptions of sibling relationships during middle childhood and adolescence. *Child Development, 61*, 1387–1398.

Feinberg, M. E., McHale, S. M., Crouter, A. C., & Cumsille, P. (2003). Sibling differentiation: Sibling and parent relationship trajectories in adolescence. *Child Development, 74*, 1261–1274.

Furman, W., & Buhrmester, D. (1985). Children's perceptions of the personal relationships in their social networks. *Developmental Psychology, 21*, 1016–1024.

Helwig, C. (2006). The development of personal autonomy throughout cultures. *Cognitive Development, 21*, 458–473.

Howe, N., Rinaldi, C., Jennings, M., & Petrakos, H. (2002). ''No! The lambs can stay out because they got cozies'': Constructive and destructive sibling conflict, pretend play, and social understanding. *Child Development, 73*, 1360–1473.

Kim, J., McHale, S. M., Osgood, D., & Crouter, A. C. (2006). Longitudinal course and family correlates of sibling relationships from childhood through adolescence. *Child Development, 77*, 1746–1761.

Laursen, B., Coy, K. C., & Collins, W. A. (1998). Reconsidering changes in parent-child conflict across adolescence: A meta-analysis. *Child Development, 69*, 817–832.

McGuire, S., Manke, B., Eftekhari, A., & Dunn, J. (2000). Children's perceptions of sibling conflict during middle childhood: Issues and sibling (dis)similarity. *Social Development, 9*, 173–190.

Montemayor, R. (1983). Parents and adolescents in conflict: All families some of the time and some families most of the time. *Journal of Early Adolescence, 3*, 83–103.

Nucci, L. P., & Lee, J. (1993). Morality and personal autonomy. In G. G. Noam & T. Wren (Eds.), *The moral self: Building a better paradigm* (pp. 123–148). Cambridge, MA: MIT Press.

Robin, A. L., & Foster, S. L. (1989). *Negotiating parent-adolescent conflict: A behavioral family systems approach*. New York: Guilford.

Ross, H. S., Filyer, R. E., Lollis, S. P., Perlman, M., & Martin, J. L. (1994). Administering justice in the family. *Journal of Family Psychology, 8*, 254–273.

Smetana, J. G. (1989). Adolescents' and parents' reasoning about actual family conflict. *Child Development, 60,* 1052–1067.

Smetana, J. G. (2006). Social-cognitive domain theory: Consistencies and variations in children's moral and social judgments. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 119–153), Mahwah, NJ: Erlbaum.

Smetana, J. G., & Asquith, P. (1994). Adolescents' and parents' conceptions of parental authority and adolescent autonomy. *Child Development, 65,* 1147–1162.

Smetana, J. G., & Daddis, C. (2002). Domain-specific antecedents of parental psychological control and monitoring: The role of parenting beliefs and practices. *Child Development*, *73*, 563–580.

Child Development, March/April 2010, Volume 81, Number 2, Pages 472–479

Categorization in 3- and 4-Month-Old Infants: An Advantage of Words Over Tones

Alissa L. Ferry, Susan J. Hespos, and Sandra R. Waxman
Northwestern University

Neonates prefer human speech to other nonlinguistic auditory stimuli. However, it remains an open question whether there are any conceptual consequences of words on object categorization in infants younger than 6 months. The current study examined the influence of words and tones on object categorization in forty-six 3- to 4-month-old infants. Infants were familiarized to different exemplars of a category accompanied by either a labeling phrase or a tone sequence. In test, infants viewed novel category and new within-category exemplars. Infants who heard labeling phrases provided evidence of categorization at test while infants who heard tone sequences did not, suggesting that infants as young as 3 months of age treat words and tones differently vis-à-vis object categorization.

Questions concerning the relation between linguistic and conceptual organization hold a central position in the cognitive sciences. Most of the research investigating the dawning of this relation has focused on early word learning and has documented that by roughly children's first birthdays, words support the formation of object categories (e.g., Balaban & Waxman, 1997; Fulkerson & Haaf, 2003; Namy, 2001; Waxman & Booth, 2003; Waxman & Braun, 2005; Woodward & Hoyne, 1999). More recent work has identified a link between words and object categories in infants as young as 6 months (Fulkerson & Waxman, 2007). Our goal was to consider this phenomenon in infants at 3 and 4 months.

Background

Waxman and Markow (1995) showed that for infants as young as 12 months of age, providing the same name for a set of distinct objects (e.g., four different animals) highlights commonalities among them and supports the formation of an object category (e.g., animal). For half of the infants in their task, the objects (e.g., the four animals) were introduced in conjunction with a novel noun (e.g., ''See the *fauna*?''); for others, the very same objects were introduced with a general orienting phrase containing no novel word (e.g., ''See here?''). At test, infants viewed two objects: a new member of the now-familiar object category (e.g., another animal) and a member of a novel object category (e.g., a vehicle), and heard, ''See what I have?'' Infants in the word condition showed a reliable preference for the novel object (e.g., a vehicle); those in the no word control condition revealed no preference. These outcomes document an early link between words and object categories in infants just beginning to build a lexicon.

Additional evidence documents the specificity of this phenomenon within the 1st year of life. At 13–14 months, infants distinguish between *kinds* of novel words, mapping novel words presented as nouns (but not adjectives) specifically to object categories (Booth & Waxman, 2009; Waxman & Booth, 2001). But younger infants' expectations about words are less precise. At 11 months, they do not distinguish between kinds of words (nouns and adjectives) in categorization tasks (Waxman & Booth, 2003). Nonetheless, their expectations are not so general as to include novel words under any circumstance: If the *same* novel word is applied to a set of distinct objects, 12-month-olds successfully

---

This research was supported by a grant from National Institute of Health (HD030410 to S.R.W.) and the National Science Foundation (0718513 to S.J.H.). We are indebted to the parents who kindly agreed to have their infants participate in the experiment and to the member of the Infant Cognition Lab who helped to collect the data.

Correspondence concerning this article should be addressed to Alissa L. Ferry, Department of Psychology, Northwestern University, 2029 Sheridan Road, Evanston, IL 60208. Electronic mail may be sent to alissaferry@u.northwestern.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8102-0005

form an object category, but if a *different* word is applied to each, they do not (Waxman & Braun, 2005). This indicates that by 12 months, infants successfully cull novel words from the ongoing speech stream, detect whether the same or different words have been applied to a set of objects, and expect that objects named by the same word (but not those named by different words) are members of a shared category.

Importantly, although other kinds of auditory stimuli, including tones and melodic sequences, are engaging to infants, these stimuli do not promote object categorization even when they are consistently applied. Fulkerson and Waxman (2007) compared the influence of novel words and tone sequences on object categorization in infants at 6 and 12 months. Infants viewed pictures of exemplars from the same category (e.g., dinosaurs), presented sequentially. Infants in the word condition heard a naming phrase as they viewed each picture (e.g., "Look at the toma! Do you see the toma?"); those in the tone condition heard a sequence of pure tones, matched precisely to the word condition for timing, duration, and volume. At test, all infants viewed pictures of a new exemplar from the familiar object category (e.g., a new dinosaur) and an exemplar from a novel category (e.g., a fish), presented side by side. Infants in the word condition revealed a reliable preference for the member of the novel category, indicating that they successfully formed the object category. Infants in the tone condition failed to reveal reliable preferences, suggesting that they had not formed an object category. Thus, by 6 months of age, words facilitate object categorization in a way that well-matched tone sequences do not.

It is important to point out that the claim is not that object categorization depends entirely on the presentation of words. Clearly, there are many circumstances in which humans, and even nonhuman animals, categorize successfully even in the absence of a novel word (e.g., Hespos & Spelke, 2004; Mareschal & Quinn, 2001; Plunkett, Hu, & Cohen, 2008; Quinn, 2002; Santos, Sulkowski, Spaepen, & Hauser, 2002). The claim is more measured and more precise: Even before infants produce words on their own, naming a set of objects with a consistently applied word promotes the formation of object categories, highlighting commonalities among objects in a way that other well-matched auditory stimuli (e.g., tone sequences) do not (Brown, 1957; Waxman & Markow, 1995). Interpreted in this light, the claim is that words support the ubiquitous process of categorization and are especially effective in highlighting commonalities among objects that may otherwise have gone undetected (Fulkerson & Waxman, 2007; Gelman & Kalish, 2006; Plunkett et al., 2008; Waxman & Markow, 1995).

In the current study, we sought to examine the precursors to this phenomenon in infants at 3 and 4 months of age. We focused our efforts here because although 3- and 4-month-old infants share with older infants an interest in human speech, it is not until roughly 6 months of age that they begin to parse individual words (especially their own names) from the stream of fluent speech (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005; Jusczyk & Aslin, 1995). Apparently, then, infants' treatment of words may differ importantly from 3 to 6 months. Although infants prefer human speech to many types of nonlinguistic auditory stimuli even at birth (Ramus, Hauser, Miller, Morris, & Mehler, 2000; Vouloumanos & Werker, 2007), whether this early preference has any consequences on categorization behaviors in infants younger than 6 months remains very much an open question.

Our goal was to examine object categorization under two different auditory conditions, one involving human speech (the word condition) and the other serving as a nonlanguage control (the tone condition). This design permitted us to discover whether words and tones exert the same influence on infants' categorization, or whether words enjoy a privileged status vis-à-vis categorization as early as 3 or 4 months of age.

Method

Participants

Participants included 46 healthy, full-term infants divided into two age groups. There were twenty-four 4-month-olds (9 males, 15 females) with an average age of 4 months 5 days (range = 3 months 23 days to 4 months 21 days) and twenty-two 3-month-olds (13 males, 9 females) with an average age of 3 months 7 days (range = 2 months 19 days to 3 months 21 days). Half of the infants within each age group were randomly assigned to the word condition and the others to the tone condition. All infants in the language condition were from families where English was the predominant language spoken in the home. An additional 13 infants were tested but excluded due to parental or sibling interference (2), fussiness (10), or falling asleep during the study (1). An additional 17 infants, who looked at the stage for less than half of the familiarization period, were included in separate analyses (5 were 3-month-old infants and 12 were 4-month-old infants). Other studies have shown that the amount of exposure infants receive during the familiarization period can affect whether infants will show a novelty or a familiarity preference during test trials (e.g., Theissen, Hill, & Saffron, 2005). We suspected that the low exposure to the familiarization stimuli may have impacted their ability to form a category and separate post hoc analyses were conducted to examine that possibility.

Participant families were recruited from a commercial mailing list from the surrounding area and contacted by mail and telephone. They were given a t-shirt or a book as a gift but were not otherwise compensated for their participation. The participants were predominantly from White, middle-class families. The ethnicity of the sample was 80% non-Hispanic. The racial make-up was 75% White, 5% Asian, 3% Black/African American, 9% multiracial, and 8% chose not to answer.

Stimuli

The stimuli were identical to the stimuli used in Fulkerson and Waxman (2007).

Visual stimuli. Twenty slides of line-drawn dinosaurs and fish were produced to form two familiarization sets of eight stimuli each and two sets of two test stimuli each (see Figure 1). Stimuli were outlined in black and filled in with a solid color. Within each familiarization set, stimuli varied in color; within each test set, stimuli were matched in color.

Auditory stimuli. Two sets of naming phrases (“Look at the *toma/modi.* Do you see the *toma/modi?*”) were spoken by a female in the infant-directed speech register and recorded for presentation. Two sequences of pure tones (400 and 800 Hz), matched to the naming phrases in duration (2.2 s), pause length, and volume (82 dB) were created.

Apparatus

Infants sat on a parent’s lap in a small room in front of a wooden puppet stage measuring 243.5 cm high, 128 cm wide, and 61 cm deep. The infants faced an opening in the front of the stage (93 cm above the floor, 61 cm high, and 106 cm wide). Visual stimuli were projected onto a white screen via a slide projector from behind the stage. A white projection screen was on the stage behind a sheet of black foam core. There were two 18 × 18 cm windows cut into the cardboard. There was 11 cm of separation between the two windows, which were centered approximately 75 cm from the infants’ eye level. Auditory stimuli were presented through a central speaker hidden 56 cm below of the screen. Infants were videotaped through a 3-cm hole in the stage located 12 cm below the projection screens.

Procedure

The novelty-preference task included a familiarization phase and a test phase (see Figure 1). All infants saw the same visual stimuli materials (either dinosaurs or fish). What differed was the accompanying audio: Infants were randomly assigned to either the word or the tone condition. Parents were instructed not to influence their infant’s attention in any way and to close their eyes during test trials.

Familiarization phase. Infants were presented with a sequence of eight different exemplars from a single category (either dinosaurs or fish). The category used was counterbalanced across both word and tone conditions. In the word condition, infants heard a naming phrase as they viewed the stimuli; in the tone condition, infants heard a sequence of pure tones. In both conditions, the auditory stimulus was introduced once at the onset of the visual stimulus and again 10 s after the onset of the visual stimulus. The trials were a fixed duration of 20 s, and the order of presentation was randomized across participants. The position of the image alternated sides from trial to trial. The following factors were counterbalanced across participants: left/right position of the first image, use of toma/modi or 400/800 Hz, and familiarization to a picture of a dinosaur or a fish.

Test phase. All infants saw the same two test stimuli: a new exemplar from the familiar category and an exemplar from a novel category. These were presented side by side in silence. The test phase began with the infants’ first visual fixation to one of the stimuli and lasted until the infant had accumulated 10 s of looking to the test stimuli. The left–right positioning of the test stimuli was counterbalanced across infants.

Coding. Trained observers, blind to the condition assignment, coded infants’ visual fixations online. Later, off-line frame-by-frame coding to determine left and right looks was conducted using ELAN (2007). Thirty-three percent of infants in each age and condition were recoded by an

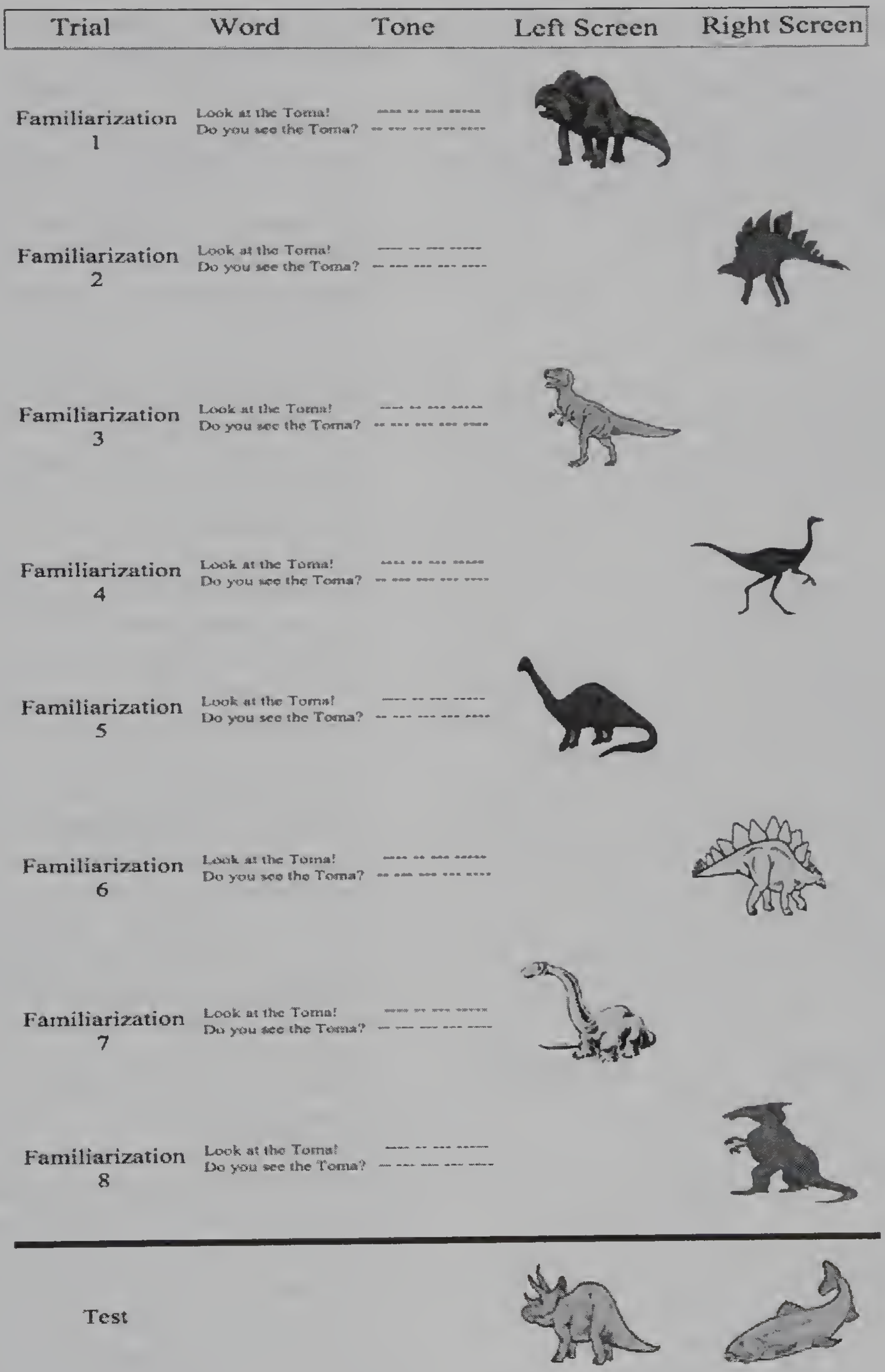

Trial
Word
Tone
Left Screen
Right Screen
Familiarization 1
Look at the Toma! Do you see the Toma?
Familiarization 2
Look at the Toma! Do you see the Toma?
Familiarization 3
Look at the Toma! Do you see the Toma?
Familiarization 4
Look at the Toma! Do you see the Toma?
Familiarization 5
Look at the Toma! Do you see the Toma?
Familiarization 6
Look at the Toma! Do you see the Toma?
Familiarization 7
Look at the Toma! Do you see the Toma?
Familiarization 8
Look at the Toma! Do you see the Toma?
Test

Figure 1. A representation of the procedure and stimuli for both the word and tone conditions.
Note. In familiarization trials, infants saw eight different line drawings of dinosaurs (or fish). Each dinosaur was accompanied by either a voice saying "Look at the *toma*! Do you see the *toma*?" (word condition) or a series of pure tones (tone condition). During test trials, infants saw a dinosaur and a fish and coders measured how long each infant looked at each image.

independent observer. Reliability between observers was 91% during familiarization and 90% during test.

To measure categorization, a preference score was calculated for each infant by dividing their looking time to the novel-category exemplar at test by their total looking time (10 s). Chance performance was .50. Six 3-month-olds (three in each condition) looked exclusively to one test stimulus. Because we were uncertain whether these infants noticed the other test stimulus, we presented a second test trial and used their performance on this trial to calculate a preference score. We obtain the same results (see below) whether we include these preference scores for these infants or exclude their data entirely.

Preliminary analyses revealed no significant effect of sex, the specific word or tone used during visual stimuli, or left or right position of the test stimuli; the data were therefore collapsed across these variables in subsequent analyses.

Results

The results, depicted in Figure 2, reveal that words and tones have distinct effects on object categorization. Importantly, infants hearing words and tones accumulated comparable looking durations during familiarization, yet only those hearing words revealed evidence of categorization at test.

An analysis of performance during familiarization revealed no differences in infants' looking times in the word condition (*M* = 114.51, *SD* = 20.85) and tone condition (*M* = 118.28, *SD* = 22.28). A univariate analysis of variance (ANOVA) with total looking time during familiarization as a dependent variable and age (3 or 4 months) and conditions (word or tone) as between-subject variables confirmed that there were no significant main effects, $Fs(1, 42) < 1$, or interaction, $F(1, 42) = 2.94$, $p = .09$. This suggests that the words and tone sequences presented in this experiment were equally engaging to our young participants. Importantly however, differences between these conditions did emerge at test.

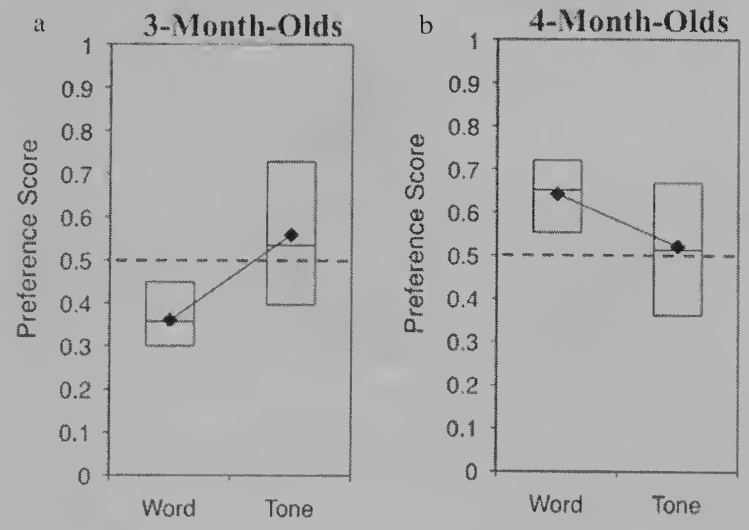


Figure 2. Box plots for mean looking times during test trials for each condition. (a) Preference score for 3-month-olds in the word and tone conditions. (b) Preference score for the 4-month-olds in the word and tone conditions.

Note. Preference score is the looking time to the novel-category exemplar divided by the total looking time. The diamonds represent the mean, the central line in the box is the median, and the upper and lower portions of the box represent the first and third quartiles on either side of the mean. The dotted line represents a chance score.

An Age (3 or 4 months) × Condition (word or tone) ANOVA using infants' preference scores at test as a dependent variable revealed no main effects, but it did reveal a significant Age × Condition interaction, $F(1, 42) = 6.08$, $p = .018$, $\eta^2 = .126$. Further analyses revealed that 3- and 4-month-old infants performed comparably in the tone condition, $F(1, 22) = 0.08$, $p = .78$; at neither age did performance with tones differ from the chance level: 3-month-olds ($M = 0.56$, $SD = 0.24$), $t(11) = .84$, $p = .42$, and 4-month-olds ($M = 0.53$, $SD = 0.22$), $t(11) = .50$, $p = .63$. In contrast, infants' performance in the word condition varied reliably as a function of age, $F(1, 20) = 12.40$, $p = .002$, $\eta^2 = .38$: Three-month-olds showed a reliable preference for the familiar test object ($M = 0.36$, $SD = 0.14$), $t(9) = -3.14$, $p = .01$, and 4-month-olds showed a reliable preference for the novel ($M = 0.64$, $SD = 0.22$), $t(11) = 2.27$, $p = .04$.

An examination of individual infants' performance echoed this pattern of results. We tabulated the number of infants at each age and in each condition demonstrating either a novelty or a familiarity preference and compared these to the number expected by the binomial distribution. In the word condition, nine of the ten 3-month-old infants showed a familiarity preference, and ten of the twelve 4-month-old infants showed a novelty preference (both *p* values < .05). In the tone condition, five of the twelve 3-month-old infants showed a familiarity preference, and seven of the twelve 4-month-old infants (both *p* values > .7) showed a novelty preference.

These results reveal that by 3–4 months of age, presenting infants with words influences object categorization in a way that presenting them with tones does not. Moreover, this influence is expressed as a familiarity preference at 3 months but as a novelty preference at 4 months. To better understand this shift from familiarity to novelty preferences, we examined the correlation between performance in this task and age (in days). For infants in the word condition, there was a strong positive correlation between age and performance, $r(22) = .61$, $p = .003$ (see Figure 3). For those in the tone condition, there was no corresponding correlation, $r(24) = -.18$, $p = .41$.

This developmental shift from familiarity to novelty preference, although unanticipated in the context of the current experiment, is not unprecedented in the infant categorization literature. Both novelty and familiarity preferences are indices of categorization. Roder, Bushnell, and Sasseville

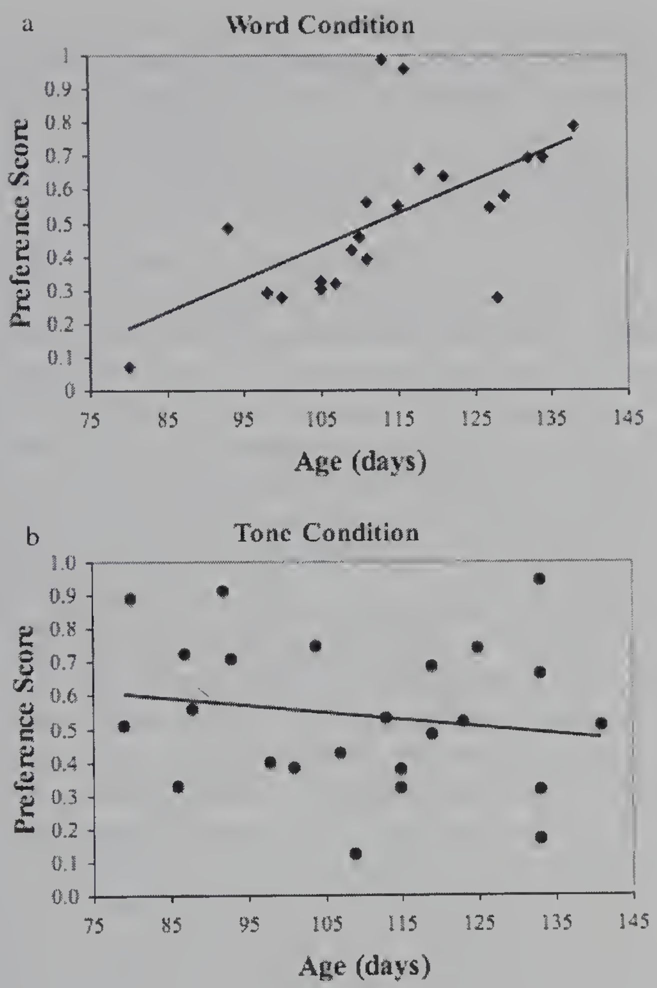


Figure 3. Scatter plot depicting the correlation between each infant's age (in days) and their preference scores. (a) Word condition. (b) Tone condition.
Note. The lines represent the best fit of the correlation between age and preference score within each condition.

(2000) found that this shift in preference was due to the amount of exposure infants had accumulated; infants first demonstrated a reliable familiarity preference prior to expressing a novelty preference. Interpreted in this light, the current results suggest that for 3-month-olds in the word condition, the process of categorization is indeed underway (after all, their performance differed significantly from chance) but that they may require more exposure before they shift from a familiarity to a novelty preference (Roder et al., 2000; Rose, Gottfried, Melloy-Carminar, & Bridger, 1982; Wagner & Sakovits, 1986). If this interpretation is correct, then perhaps 4-month-olds in the word condition who happen to accrue little looking time during familiarization materials should reveal a familiarity preference as well.

To test this possibility, we conducted a post hoc analysis on the twelve 4-month-old infants (five in the word condition, seven in the tone condition) who had been eliminated from the previous analyses because they accrued too little exposure to the familiarization images (recall that for the previous analyses, we excluded infants who looked for less than 50% of the familiarization trials). Performance in the tone condition was unaffected by exposure. These tone condition infants, like those in the main analyses, did not differ significantly from chance, (M = 0.42, SD = 0.22), $t(6) = -.99$, $p = .36$. In contrast, 4-month-olds in the word condition showed a significant familiarity preference (M = 0.35, SD = 0.07), $t(4) = -3.72$, $p = .008$. This is consistent with our interpretation of the shift from a familiarity to a novelty preference at 3 and 4 months, respectively: For the stimuli presented here, infants' expression of a familiarity versus novelty preference at test is related to their exposure during familiarization.

Discussion

The results demonstrate that words influence object categorization at 3 and 4 months, and do so in a way that is distinct from the influence of carefully matched tones. This study makes two primary contributions. First, it offers the earliest evidence to date for a link between words and object categories. While previous studies have demonstrated this link in infants as young as 6 months (Fulkerson & Waxman, 2007), the current findings offer insight into the origins of this ability in 3- and 4-month-old infants. Second, these results indicate that even at this early developmental moment, words exert an influence on object categorization that is distinct from that of tones (Balaban & Waxman, 1997; Fulkerson & Haaf, 2003; Fulkerson & Waxman, 2007; Waxman & Booth, 2003; Waxman & Braun, 2005; Waxman & Markow, 1995; Woodward & Hoyne, 1999; Xu, 1999, 2002).

This finding, intriguing in and of itself, also calls into question the claim that the facilitative effect of words can be accounted for entirely by appealing to matters of paired associations. In the current experiment, the tones as well as the words were correlated perfectly with the visually presented objects. If the categorization effect could be attributed to nothing more than an association between auditory and visual stimuli (e.g., Robinson & Sloutsky, 2007), then infants should have performed comparably, whether they heard words or tones. The fact that they did not serves as evidence that words exert an influence that cannot be accounted for entirely by association, even in infants as young as 3 months of age.

The current results also make important contact with evidence of infants' categorization behaviors.

The classic behavioral index of category formation in infants rests on their treatment of novel versus familiar test stimuli. Previous evidence concerning the link between words and object categorization in infants at 6, 9, and 12 months of age has consistently revealed a preference for the novel test stimuli (Balaban & Waxman, 1997; Fulkerson & Waxman, 2007). In the experiment reported here, 4-month-olds display this novelty preference, but 3-month-olds reveal a familiarity preference. As we have pointed out, familiarity and novelty preferences are both indices of object categorization (Aslin, 2007; Roder et al., 2000). For example, Roder et al. (2000) demonstrated that infants tend to exhibit a familiarity preference early in processing followed by a novelty preference in subsequent trials. The mechanism underlying the expression of novelty versus familiarity preferences is not well understood, but several hypotheses have been advanced. For example, some have argued that familiarity preferences are the result of incomplete encoding (Hunter & Ames, 1988; Roder et al., 2000; Rose et al., 1982; Wagner & Sakovits, 1986) or that subcomponents of fully encoded stimuli can trigger recognition responses for familiarity that are stronger than attention response to novelty (Kaplan & Werner, 1987). Others have suggested that factors including maturity, processing speed, and encoding efficiency are involved (Colombo, 2002; Johnson, Slemmer, & Amso, 2004). The absence of an independent assessment for these confounded cognitive processes makes it impossible to predict familiarity or novelty effects, although the consensus among infancy researchers is that both effects are behavioral indices of category formation. Future studies may pursue this issue by systematically varying exposure to the words, the visual stimuli, or to both over the course of familiarization.

The current findings also raise questions about the range of auditory stimuli that promote categorization at 3 and 4 months of age. We have documented that novel words (presented as nouns) promote categorization in a way that tones do not. In future work, it will be important to identify which auditory features, present in the word condition but absent in the tone condition, might underlie this effect. It will also be important to ascertain which other complex sounds (e.g., backward speech, filtered speech) also promote categorization in young infants and whether this facilitative effect is specific to human vocalizations.

It will also be important to understand better the mechanism underlying the facilitative effect of words on categorization at 3 and 4 months. As we have pointed out, it is unlikely that infants in the current experiment parsed the novel words from the labeling phrases in which they were embedded; it is also unlikely that they somehow understood the meaning of the words and phrases we presented. Why, then, did words promote categorization over and above the effect of tone sequences? We speculate that in the context of hearing human speech, very young infants become engaged in attentional and behavioral processes that they might otherwise not engage in. An analogy may illustrate this point. Recent research has documented a dramatic influence of vocalizations on the behavior of Nile crocodile hatchlings and mothers. Under naturalistic conditions, the vocalizations made by hatchlings just before they hatch can be heard by mothers and by hatchlings in neighboring eggs in the nest. In an ingenious experiment, Vergne and Mathevon (2008) presented hatchlings and mothers with recordings of the vocalizations emitted by other hatchlings just before they had hatched. These recordings had very specific behavioral effects; they motivated hatchlings to tap on their shells and mothers to open their nests. Obviously, it is unlikely that these vocalizations exerted their effects because of the "meaning" they conveyed to mothers and other hatchlings. Instead, exposure to these vocalizations appears to have engendered a series of behaviors that happen to promote hatching and nesting. How does this analogy bear on infants' performance in the current experiment? We suspect that human speech, and perhaps especially infant-directed speech, engenders in young infants a kind of attention to the surrounding objects that promotes categorization. We proposed that over time, this general attentional effect will become more refined, as infants begin to cull individual words from fluent speech, to distinguish among individual words and kinds of words, and to map those words to meaning. A goal for future research will be to identify the processes underlying this developmental trajectory.

References

Aslin, R. N. (2007). What's in a look? *Developmental Science*, *10*, 48–53.

Balaban, M. T., & Waxman, S. R. (1997). Do words facilitate object categorization in 9-month-old infants? *Journal of Experimental Child Psychology*, *64*, 3–26.

Booth, A. E., & Waxman, S. R. (2009). A horse of a different color: Specifying with precision infants' mapping of novel nouns and adjectives. *Child Development*, *80*, 15–22.

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech-stream segmentation. *Psychological Science, 16*, 298–304.

Brown, R. W. (1957). Linguistic determinism and the part of speech. *Journal of Abnormal & Social Psychology, 55*, 1–5.

Colombo, J. (2002). Infant attention grows up: The emergence of a developmental cognitive neuroscience perspective. *Current Directions in Psychological Science, 11*, 196–200.

ELAN. (2007). *ELAN*. Nijmegen, Netherlands: Language Archiving Technologies.

Fulkerson, A. L., & Haaf, R. A. (2003). The influence of labels, non-labeling sounds, and source of auditory input on 9- and 15-month-olds' object categorization. *Infancy, 4*, 349–369.

Fulkerson, A. L., & Waxman, S. R. (2007). Words (but not tones) facilitate object categorization: Evidence from 6- and 12-month-olds. *Cognition, 150*, 218–228.

Gelman, S. A., & Kalish, C. W. (2006). Conceptual development. In D. Kuhn & R. Siegler (Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (pp. 687–733). New York: Wiley.

Hespos, S. J., & Spelke, E. S. (2004). Conceptual precursors to language. *Nature, 430*(6998), 453–456.

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. In L. P. Lipsitt (Ed.), *Advances in child development and behavior* (pp. 69–95). New York: Academic Press.

Johnson, S. P., Slemmer, J. A., & Amso, D. (2004). Where infants look determines how they see: Eye movements and object perception performance in 3-month-old infants. *Infancy, 6*, 185–201.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology, 29*, 1–23.

Kaplan, P. S., & Werner, J. (1987). Habituation, response to novelty, and dishabituation in human infants: Tests of a dual-process theory of visual attention. *Journal of Experimental Child Psychology, 14*, 83–109.

Mareschal, D., & Quinn, P. C. (2001). Categorization in infancy. *Trends in Cognitive Sciences, 4*, 443–450.

Namy, L. L. (2001). What's in a name when it isn't a word? 17-month-olds' mapping of non-verbal symbols to object categories. *Infancy, 2*, 73–86.

Plunkett, K., Hu, J., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition, 106*, 665–681.

Quinn, P. C. (2002). Category representation in young infants. *Current Directions in Psychological Science, 11*, 66–70.

Ramus, F., Hauser, M. D., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science, 288*(5464), 349–351.

Robinson, C. W., & Sloutsky, V. M. (2007). Linguistic labels and categorization in infancy: Do labels facilitate or hinder? *Infancy, 11*, 223–253.

Roder, B. J., Bushnell, E. W., & Sasseville, A. M. (2000). Infants' preferences for familiarity and novelty during the course of visual processing. *Infancy, 1*, 491–507.

Rose, S. A., Gottfried, A. W., Melloy-Carminar, P., & Bridger, W. H. (1982). Familiarity and novelty preferences in infant recognition memory: Implications for information processing. *Developmental Psychology, 18*, 704–713.

Santos, L. R., Sulkowski, G. M., Spaepen, G. M., & Hauser, M. D. (2002). Object individuation using property/kind information in rhesus macaques (*Macaca mulatta*). *Cognition, 83*, 241–264.

Thiessen, E. D., Hill, E., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy, 7*, 53–71.

Vergne, A. L., & Mathevon, N. (2008). Crocodile egg sounds signal hatching time. *Current Biology, 18*(12), R513–R514.

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science, 10*, 159–164.

Wagner, S. H., & Sakovits, L. J. (1986). A process analysis of infant visual and cross-modal recognition memory. In L. P. Lipsitt (Ed.), *Advances in infancy research* (Vol. 4, pp. 195–217). Norwood, NJ: Ablex.

Waxman, S. R., & Booth, A. E. (2001). Seeing pink elephants: Fourteen-month-olds' interpretations of novel nouns and adjectives. *Cognitive Psychology, 43*, 217–242.

Waxman, S. R., & Booth, A. E. (2003). The origins and evolution of links between word learning and conceptual organization: New evidence from 11-month-olds. *Developmental Science, 6*, 128–135.

Waxman, S. R., & Braun, I. (2005). Consistent (but not variable) names as invitations to form object categories: New evidence from 12-month-old infants. *Cognition, 95*, B59–B68.

Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology, 29*, 257–302.

Woodward, A. L., & Hoyne, K. L. (1999). Infants' learning about words and sounds in relation to objects. *Child Development, 70*, 65–77.

Xu, F. (1999). Object individuation and object identity in infancy: The role of spatiotemporal information, object property information, and language. *Acta Psychologica, 102*, 113–136.

Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition, 85*, 223–250.

Child Development, March/April 2010, Volume 81, Number 2, Pages 480–486

The Complex Relation Between Bullying, Victimization, Acceptance, and Rejection: Giving Special Attention to Status, Affection, and Sex Differences

René Veenstra, Siegwart Lindenberg, Anke Munniksma, and Jan Kornelis Dijkstra
University of Groningen

To understand the complex nature of bullies' acceptance and rejection, this article considered goal-framing effects of status and affection as they relate to the gender of the bully (male vs. female bullies), the target (male vs. female victims), and the evaluator (acceptance and rejection from male vs. female classmates). The hypotheses were tested with data from a social network questionnaire conducted in 26 elementary school classes (N = 481 children; M_{age} = 10.5 years). The findings revealed that bullies were only rejected by those for whom they were a potential threat and that bullies generally chose their victims so as to minimize loss of affection by choosing victims who were not likely to be defended by significant others.

Bullying is a common phenomenon in school classes. In this study, we try to clarify the complex relations of bullying with acceptance and rejection. Because peer processes, especially in childhood and early adolescence, often show gender segmentation, we are interested in whether bullying toward same-gender classmates has the same effect on peer acceptance and rejection as bullying toward other-gender classmates. We argue that besides taking the gender of the bully (male vs. female bullies) into account, we should also take the gender of the target (male vs. female victims) and the evaluator (acceptance and rejection by male vs. female classmates) into account to explain the relation between bullying, victimization, and peer status. The following questions will be examined: How are bullying and victimization related to peer status among same-gender and other-gender classmates? Is bullying toward same-gender classmates differently related to status than bullying toward other-gender classmates? And how is this for victimization by same-gender or other-gender classmates?

Theoretical Elaboration

In the goal-framing approach, focal goals are hypothesized to influence what people attend to, what knowledge is being activated, how people evaluate things, and how they process information (Lindenberg, 2006). People are aware of aspects in the situation that potentially help or hinder their goal pursuit, and they positively evaluate (like) the former and negatively evaluate (dislike) the latter. Liking and disliking are thus the result of different goal-related processes. This goal-framing approach has recently been applied to questions of acceptance and rejection (Dijkstra, Lindenberg, & Veenstra, 2007), to questions concerning who bullies whom (Veenstra et al., 2007), and to the role of popular adolescents in bullying (Dijkstra, Lindenberg, & Veenstra, 2008).

When studying interactions among children, what goals should be considered? Status and affection goals have frequently been identified as important for all human beings (Barkow, 1989; Baumeister & Leary, 1995; Huberman, Loch, & Önçüler, 2004; Ormel, Lindenberg, Steverink, & Vonkorff, 1997). Although we do not measure these goals directly, we have good evidence for their importance. Pendell (2002) has reviewed much literature that shows affection to be a universal need. The evolutionary and developmental importance of affection has also been shown (MacDonald, 1992). Status has also been established as a universal goal (Barkow, 1989; Huberman et al., 2004), and the importance of this goal for bullying has recently been directly assessed (see Sijtsema, Veenstra, Lindenberg, & Salmivalli, 2009). Both goals are prominent in childhood and preadolescence as well

---

Correspondence concerning this article should be addressed to René Veenstra, Sociology, University of Groningen, Grote Rozenstraat 31, Groningen 9712 TG, the Netherlands. Electronic mail may be sent to d.r.veenstra@rug.nl.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8102-0006

(Buhrmester, 1990; Hawley, 2003; Jarvinen & Nicholls, 1996; Oldehinkel, Rosmalen, Veenstra, Dijkstra, & Ormel, 2007). Thus, it seems to be a safe assumption that bullies, like other human beings, want to realize status *and* affection.

Given these goals, the crucial aspect of the pursuit of status is that it is conditioned by the pursuit of affection. People want both status and affection and are not eager to forego one for the sake of the other (Lindenberg, 2001). For bullying this means that children who want to dominate will be keenly aware of opportunities to do so without risking loss of affection from significant others (O'Connell, Pepler, & Craig, 1999). Thus, bullies (i.e., instrumental and not reactive bullies) can be expected to strategically choose victims who are already rejected (disliked) by significant others in the class. Bullies are likely to divide the classroom into potential sources of affection (significant others) and potential sources of domination (victims for whom the significant others do not care). For children, the significant others are likely to be same-gender classmates (Dijkstra et al., 2007; Maccoby, 1998; Martin & Halverson, 1981). Thus, bullies can also be expected to desire to be accepted by the significant same-gender classmates and not to care much about rejection by the rest (Olthof & Goossens, 2008).

Bullies thwart the goal pursuit of victims and, on the basis of goal-framing theory, can be predicted to be rejected by their victims and by others for whose goal pursuit they are a potential threat. From this follows our first set of hypotheses (on selective threat and rejection). We expect that the rejection of bullies will only come from the gender to which the victim belongs: (1a) Bullying same-gender classmates is related to being rejected by same-gender classmates only. (1b) Bullying other-gender classmates is related to being rejected by other-gender classmates only.

With regard to affection, goal-framing theory predicts that male bullies are likely to strategically choose victims who pose a minimal risk for a loss of affection: those boys who have low acceptance among boys (i.e., they are not important for affection) and are rejected by boys. Considering that children rarely have best friends in the other-gender group, we argue that for boys it is slightly different to bully girls: Male bullies are likely to choose those girls as victims who are rejected by boys (liking is not an issue here). For female bullies, we expect the converse: Female bullies are likely to choose those boys as victims who are rejected by girls. There is no a priori reason to assume that for them the two goals work differently, although there will be fewer girls for whom domination is a prominent goal (Espelage, Mebane, & Adams, 2004; Hanish & Guerra, 2004).

From the foregoing, we can deduce our second set of hypotheses (on bullying and the avoidance of loss of affection). We expect that bullies will avoid loss of affection by choosing victims that are not cared for by significant others: (2a) If children bully same-gender classmates, they will focus on those potential victims who are rejected by and have low acceptance from the bullies' same-gender classmates. (2b) If children bully other-gender classmates, they will focus on those potential victims who are rejected by the bullies' same-gender classmates. (2c) There will be no negative relation between being a bully and the level of acceptance from boys and girls (bullies do not lose affection). If all these hypotheses are supported by the data, it will be likely that it is indeed avoidance of loss of affection that drives the results.

Note that the hypotheses contain different sources of acceptance and rejection. In theory, victims have low acceptance and are already rejected *before* being bullied. Choosing these victims is part of the bully's strategy to avoid loss of affection, in particular from same-gender classmates. Bullies are rejected if they pose a threat. Given the gender segmentation of children, they are expected to be rejected by members of the gender of the victim and ignored by the rest.

Method

Sample

Network questions on bullying and victimization were collected in middle and late elementary education (Grades 5–8 in the Netherlands). The sample yielded 481 children from 26 classrooms (23 for regular and 3 for special education): 218 girls (45.3%) and 263 boys (54.7%), with a mean age of 10.5 years ($SD = 1.5$). The mean class size was 19.4 children ($SD = 4.4$). Schools were situated in both rural and (sub)urban areas. The percentage of children with parents with a low educational level, at maximum a certificate of secondary vocational education, was 16.9%. The percentage of children from ethnic minorities (of whom at least one parent was born outside the Netherlands) was 18.7%.

Procedure

Data were collected from October 2005 to March 2007. After parental consent was obtained, children filled out the questionnaires in the school class, under the supervision of a well-trained research assistant. Ninety-seven percent of all children participated in the study. The dyadic peer-nomination assessment took place at school. The number of nominations the children could make was unlimited (they were not required to nominate anyone), and same-gender as well as other-gender nominations were allowed.

Measures

Peer acceptance and rejection. The numbers of nominations children received individually from their same- and other-gender classmates with regard to "best friends" and "dislike" were used to create measures of same- and other-gender peer acceptance and peer rejection. After the numbers of received nominations had been summed, proportions were calculated to take differences in the number of respondents per class into account, yielding scores from 0 to 1 (see Veenstra et al., 2007, for more information on this dyadic peer nomination procedure).

Bullying and victimization. The term *bullying* was defined to the students in the way formulated in the Olweus's (1996) Bully/Victim Questionnaire, which emphasizes the repetitive and intentional nature of bullying and the power imbalance between the bully and the victim. Several examples covering different forms of bullying were given. In addition, examples of behaviors that should not be considered as bullying (teasing in a friendly and playful way, fighting between children of equal strength) were also provided.

The numbers of nominations children received individually from their same- and other-gender classmates with regard to different forms of bullying and victimization were used to create measures of same- and other-gender bullying and victimization. We asked "who do you bully by . . . ?" and "by whom are you bullied by . . . ?" using five forms of bullying and victimization: (a) excluding or ignoring; (b) calling names or laughing; (c) hitting, kicking, or pinching; (d) taking things; (e) throwing things. There were no clear differences in the association of the different forms of bullying and victimization with peer status. For that reason, we combined the different forms in highly reliable scales for bullying and victimization (Cronbach's alphas = .89 and .87, respectively).

Table 1
Means and Standard Deviations of Peer Status, Bullying, and Victimization for Boys (N = 263) and Girls (N = 218)

	Boys	Girls
Acceptance by boys	.393 (.203)	.110 (.121)
Acceptance by girls	.110 (.142)	.474 (.250)
Rejection by boys	.119 (.160)	.163 (.183)
Rejection by girls	.267 (.250)	.091 (.142)
Bullying toward boys	.065 (.083)	.034 (.051)
Bullying toward girls	.064 (.092)	.035 (.053)
Victimization by boys	.041 (.070)	.032 (.057)
Victimization by girls	.023 (.045)	.022 (.053)

Analyses

We tested our hypotheses with multivariate analyses using cross-sectional data. Because both acceptance and rejection deviated from normality, we conducted regression analyses with the Tobit model, which accounts for violations of normality of the dependent variables (Long, 1997; Smith & Brame, 2003; Tobin, 1953). The regression analyses included main effects of gender, bullying toward boys and toward girls, victimization from boys and from girls, and (the significant) interaction effects between gender and either bullying or victimization. The effects for girls are equal to the main effects in Tables 2 and 3, but the effects for boys are the sum of the main and interaction effects (Aiken & West, 1991). All continuous variables were standardized for the whole sample ($M = 0$, $SD = 1$). Finally, we examined whether the effects differed by age.

Results

Table 1 shows that same-gender classmates were more accepted and less rejected than other-gender classmates by both boys and girls. Furthermore,

Table 2
Multiple Regression Analyses on Peer Rejection and Its Relation to Bullying and Victimization (N = 481)

	Peer rejection by boys	Peer rejection by girls
Gender (1 = boys)	−.77 (.18)**	.94 (.19)**
Bullying toward boys	.74 (.15)**	.36 (.15)*
Bullying toward girls	−.03 (.11)	.37 (.09)**
Victimization by boys	.35 (.06)**	.15 (.11)
Victimization by girls	.16 (.09)†	.18 (.06)**
Gender × Bullying Toward Boys	−.38 (.14)**	−.33 (.19)†
Pseudo R^2	13.8%	15.9%

Note. White–Huber standard errors that adjust for clustering of individuals within classrooms are reported. Pseudo R^2 values as obtained from Tobit analyses are reported. Values are expressed as *B* (*SE*).
$^\dagger p < .10$. $^*p < .05$. $^{**}p < .01$.

Table 3

Multiple Regression Analyses on Peer Acceptance and Its Relation to Bullying and Victimization (N = 481)

	Peer acceptance by boys	Peer acceptance by girls
Gender (1 = boys)	1.64 (.14)**	−1.56 (.15)**
Bullying toward boys	−.46 (.19)*	−.28 (.14)*
Bullying toward girls	−.01 (.08)	−.09 (.08)
Victimization by boys	.12 (.15)	.08 (.15)
Victimization by girls	−.05 (.07)	−.17 (.08)*
Gender × Bullying Toward Boys	.55 (.18)**	.50 (.17)**
Gender × Victimization by Boys	−.43 (.15)**	−.26 (.18)
Pseudo R^2	20.1%	20.4%

Note. White–Huber standard errors that adjust for clustering of individuals within classrooms are reported. Pseudo R^2 values as obtained from Tobit analyses are reported. Values are expressed as *B* (*SE*).
$^{*}p < .05$. $^{**}p < .01$.

boys were more rejected than girls. Table 1 also shows that boys were more often nominated than girls as perpetrators of bullying toward both boys and girls.

Testing the Selective Threat Hypotheses

The multivariate analyses are depicted in Tables 2 and 3. Let us first turn to the selective threat hypotheses about the rejection of bullies. As hypothesized, boys bullying boys are only rejected by boys, $b = .36$, $t(480) = 2.78$, $p < .01$, and not by girls, $b = .03$, $t(480) = 0.30$, $p = .76$. If boys bully girls, they are only rejected by girls, $b = .37$, $t(480) = 4.10$, $p < .01$, and not by boys, $b = -.03$, $t(480) = -0.33$, $p = .75$.

For girls, the selective threat hypotheses are similarly supported by the data: Girls who bully girls are rejected by girls, $b = .37$, $t(480) = 4.10$, $p < .01$, and girls who bully boys are rejected by boys, $b = .74$, $t(480) = 5.10$, $p < .01$.

Testing the Avoidance of Loss of Affection Hypotheses

With regard to avoidance of loss of affection for boys, we see from Table 2 that victims of male bullies are indeed rejected by boys only, $b = .35$, $t(480) = 6.05$, $p < .01$, and victims of female bullies are rejected by girls, $b = .18$, $t(480) = 2.92$, $p < .01$.

From Table 3, we see that male victims have a low level of acceptance among boys, $b = -.31$, $t(480) = -3.91$, $p < .01$. Among girls, female as well as male victims have a low level of acceptance, $b = -.17$, $t(480) = -2.02$, $p = .04$.

On the whole, there is, as predicted, no negative relation between bullying and acceptance. Girls even seem to welcome boys who bully boys: acceptance, $b = .22$, $t(480) = 1.93$, $p = .05$. There is one exception, though: Girls who bully boys have a low level of acceptance by both genders (see Table 3), acceptance by boys, $b = -.46$, $t(480) = -2.50$, $p = .01$, and by girls, $b = -.28$, $t(480) = -1.99$, $p = .05$.

Differences Between Middle Childhood and Preadolescence

Given the age range, we controlled for age and examined whether the results differed for middle childhood and preadolescence (median split at 10.45 years). Age had no main effect on acceptance and rejection. Only the results for peer acceptance by boys showed age differences. Girls bullying boys was more strongly related to low levels of male acceptance in middle childhood, $b = -.67$, $t(480) = -3.49$, $p < .01$, than in preadolescence, $b = -.35$, $t(480) = -1.96$, $p = .05$. Bullying girls was positively related to male acceptance in middle childhood, $b = .29$, $t(480) = 3.58$, $p < .01$, but negatively in preadolescence, $b = -.17$, $t(480) = -2.19$, $p = .03$. Victims of female bullies had low levels of male acceptance in middle childhood, $b = -.23$, $t(480) = -2.62$, $p < .01$, but not in preadolescence.

Discussion

Bullies and victims are not equally disliked. To understand the complex nature of bullies and victims' acceptance and rejection for children in middle and late elementary education, we considered goal-framing effects of status and affection as they relate to the gender of the bully (male vs. female bullies), the target (male vs. female victims), and the evaluator (acceptance and rejection from male vs. female classmates).

Our first set of hypotheses dealt with bullying as a selective threat to goal pursuit. We hypothesized that bullies would only be rejected by those for whom they were a potential threat, and this bore out. Bullies, whether male of female, were indeed rejected by the gender to whom the bullying was directed but not by the gender to whom it was not directed.

Our second set of hypotheses dealt with the bullies' avoidance of loss of affection by choosing victims that were not cared for by significant others. We predicted that bullies focus strategically on those potential same-gender victims who were rejected by and had low acceptance from same-gender classmates. For potential other-gender victims we hypothesized that children would focus on those who were rejected by the bullies' same-gender classmates. We found that victims of male bullies were indeed rejected by boys only and that male bullies were never low on acceptance. Thus, as expected, boys seem to choose their victims so as to minimize loss of affection.

Girls victimized by girls were rejected and unaccepted by girls, as goal-framing theory had predicted. However, girls who bullied boys lost acceptance and were more rejected by both genders. We have no way of saying whether this latter finding is specific to the data we used or whether it can be replicated in other studies. Girls do not frequently bully boys, but if they do they might do so because they have lost strategic control through habituation of aggression and the ensuing desensitization to its consequences (Guerra, Huesmann, & Spindler, 2003). The important point of this article is that for the vast majority of bullies, the pursuit of status *and* affection seems to foster strategic control of bullying behavior: realize one without losing the other.

There were age differences in the relation of bullying and victimization with peer acceptance by boys. Bullying toward girls was positively related to male acceptance in middle childhood, but negatively in preadolescence. Furthermore, victims of female bullies had low levels of male acceptance in middle childhood, but not in preadolescence. It might be that boys perceive cross-gender bully–victim relationships less negatively when they enter preadolescence because they become more interested in girls and may see this interaction with girls as attention (Adler & Adler, 1998; Maccoby, 1998; Pellegrini & Bartini, 2001).

There was one circumstance in which bullying was significantly positively associated with acceptance (cf. Luthar & McMahon, 1996; Salmivalli, Kaukiainen, & Lagerspetz, 2000): Male bullying of same-gender classmates was positively related to female acceptance. Again, at least for some girls, boys might have already become romantically important, with bullying fitting into a prototypical mold of being male. It should be noted, however, that the female acceptance of male–male bullying did not vary by age. Thus, older girls were not more likely to welcome male–male bullying. Yet, with regard to rejection, the selective threat effect could be observed here as well: Male bullying of girls was positively related to female rejection. By distinguishing the targets' gender, we could make clear that the evaluation of bullying depends on whether the bullying is perceived as a threat. Our findings also show that avoidance of loss of affection depends on the gender of the bully and the target.

In line with Dijkstra et al. (2007), we found that acceptance and rejection are not tied to the same process. For example, male bullying was positively related to peer rejection by the gender to which the victim belonged, but not negatively to peer acceptance. We also found that the explained variance for acceptance (about 20%) was higher than the explained variance for rejection (about 15%). It is likely that this difference is due to the fact that in preadolescence gender plays a larger role in the realization of interaction goals (and thus peer acceptance) than in disturbance or threat of disturbance of goal pursuit (and thus peer rejection). As can be seen from the multivariate analyses, gender is the factor that explains acceptance and rejection the most, but acceptance to a larger extent than rejection. The goal-framing approach also throws light on the question when behavior of the other gender will be met by positive or negative evaluation and when it will be ignored (Dijkstra et al., 2007). It depends on the contribution that a particular reaction from the other gender will make for the realization of affection and status goals.

Strengths and Limitations

Our study had a number of strengths and limitations. One strength is the elaboration of the complex nature of acceptance and rejection related to bullying. Another strength is the inclusion of boys' and girls' nominations for peer status, bullying, and victimization, each with dyadic same-gender and other-gender nominations. A strong point is also the relatively large sample. We used a sample of almost 500 children, including a proportional number of boys and girls. In view of this sample size and the use of network questions, the findings can be considered rather robust.

However, it should be taken into account that a cross-sectional correlational design was used. Even though the hypotheses derived from the goal-framing approach were supported by the results, these associations are not determinant. Ultimately, the relation between same-gender and other-gender bullying, victimization, and peer status should be tested with longitudinal data. Bidirectional influences between bullying and peer relations (Cillessen & Mayeux, 2004) would fit well into the approach taken here. Classmates who are not accepted might be even less accepted when they are bullied because victimization is likely to lower their status. Being associated with them might lower one's own status and, for this very reason, may make one more vulnerable for becoming a victim oneself (Hodges, Boivin, Vitaro, & Bukowski, 1999).

In sum, the complex nature of acceptance and rejection can be traced quite well by hypotheses derived from goal-framing theory and the assumed simultaneous pursuit of status and affection. We found that bullies are not rejected in general, but only by those for whom they are a potential threat. Bullies seem to choose their victims so as to minimize loss of affection. To this end, they are likely to bully victims that are rejected by their same-gender classmates. To understand these processes it is necessary to distinguish the gender of the bully (male vs. female bullies), the gender of the target (male vs. female victims), and the gender of the evaluator (male vs. female classmates who accept and reject bullies and victims).

References

Adler, P. A., & Adler, P. (1998). *Peer power: Preadolescent culture and identity*. New Brunswick, NJ: Rutgers University Press.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Barkow, J. H. (1989). *Darwin, sex, and status: Biological approaches to mind and culture*. Toronto, Canada: University of Toronto Press.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human-motivation. *Psychological Bulletin, 117*, 497–529.

Buhrmester, D. (1990). Intimacy of friendship, interpersonal competence, and adjustment during preadolescence and adolescence. *Child Development, 61*, 1101–1111.

Cillessen, A. H. N., & Mayeux, L. (2004). From censure to reinforcement: Developmental change in the association between aggression and social status. *Child Development, 75*, 147–163.

Dijkstra, J. K., Lindenberg, S., & Veenstra, R. (2007). Same-gender and cross-gender peer acceptance and peer rejection and their relation with bullying and helping among preadolescents: A goal-framing approach. *Developmental Psychology, 43*, 1377–1389.

Dijkstra, J. K., Lindenberg, S., & Veenstra, R. (2008). Beyond the class norm: Bullying behavior of popular adolescents and its relation to peer acceptance and rejection. *Journal of Abnormal Child Psychology, 36*, 1289–1299.

Espelage, D. L., Mebane, S. E., & Adams, R. S. (2004). Empathy, caring, and bullying: Toward an understanding of complex associations. In D. L. Espelage & S. M. Swearer (Eds.), *Bullying in American schools: A social-ecological perspective on prevention and intervention* (pp. 37–61). Mahwah, NJ: Erlbaum.

Guerra, N. G., Huesmann, L. R., & Spindler, A. (2003). Community violence exposure, social cognition, and aggression among urban elementary school children. *Child Development, 74*, 1561–1576.

Hanish, L. D., & Guerra, N. G. (2004). Aggressive victims, passive victims, and bullies: Developmental continuity or developmental change? *Merrill-Palmer Quarterly, 50*, 17–38.

Hawley, P. H. (2003). Prosocial and coercive configurations of resource control in early adolescence: A case for the well-adapted Machiavellian. *Merrill-Palmer Quarterly, 49*, 279–309.

Hodges, E. V. E., Boivin, M., Vitaro, F., & Bukowski, W. M. (1999). The power of friendship: Protection against an escalating cycle of peer victimization. *Developmental Psychology, 35*, 94–101.

Huberman, B. A., Loch, C. H., & Önçüler, A. (2004). Status as a valued resource. *Social Psychology Quarterly, 67*, 103–114.

Jarvinen, D. W., & Nicholls, J. G. (1996). Adolescents' social goals, beliefs about the causes of social success, and satisfaction in peer relations. *Developmental Psychology, 32*, 435–441.

Lindenberg, S. (2001). Social rationality versus rational egoism. In J. H. Turner (Ed.), *Handbook of sociological theory* (pp. 635–668). New York: Plenum.

Lindenberg, S. (2006). Prosocial behavior, solidarity, and framing processes. In D. Fetchenhauer, A. Flache, A. P. Buunk, & S. Lindenberg (Eds.), *Solidarity and prosocial behavior. An integration of sociological and psychological perspectives* (pp. 23–43). Berlin: Springer.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage.

Luthar, S. S., & McMahon, T. J. (1996). Peer reputation among inner-city adolescents: Structure and correlates. *Journal of Research on Adolescence, 6*, 581–603.

Maccoby, E. E. (1998). *The two sexes. Growing up apart, coming together*. Cambridge, MA: Belknap.

MacDonald, K. (1992). Warmth as a developmental construct: An evolutionary analysis. *Child Development, 63*, 753–773.

Martin, C. L., & Halverson, C. F. (1981). A schematic processing model of sex typing and stereotyping in children. *Child Development, 52*, 1119–1134.

O'Connell, P., Pepler, D., & Craig, W. (1999). Peer involvement in bullying: Insights and challenges for intervention. *Journal of Adolescence, 22*, 437–452.

Oldehinkel, A. J., Rosmalen, J. G. M., Veenstra, R., Dijkstra, J. K., & Ormel, J. (2007). Being admired or being liked: Classroom social status and depressive problems in early adolescent girls and boys. *Journal of Abnormal Child Psychology, 35*, 417–427.

Olthof, T., & Goossens, F. A. (2008). Bullying and the need to belong: Early adolescents' bullying-related behavior and the acceptance they desire and receive from particular classmates. *Social Development*, *17*, 24–46.

Olweus, D. (1996). *The Revised Olweus Bully/Victim Questionnaire*. Bergen, Norway: University of Bergen.

Ormel, J., Lindenberg, S., Steverink, N., & Vonkorff, M. (1997). Quality of life and social production functions: A framework for understanding health effects. *Social Science and Medicine*, *45*, 1051–1063.

Pellegrini, A. D., & Bartini, M. (2001). Dominance in early adolescent boys: Affiliative and aggressive dimensions and possible functions. *Merrill-Palmer Quarterly*, *47*, 142–163.

Pendell, S. D. (2002). Affection in interpersonal relationships: Not just "a fond or tender feeling." *Communication Yearbook*, *26*, 70–115.

Salmivalli, C., Kaukiainen, A., & Lagerspetz, K. (2000). Aggression and sociometric status among peers: Do gender and type of aggression matter? *Scandinavian Journal of Psychology*, *41*, 17–24.

Sijtsema, J. J., Veenstra, R., Lindenberg, S., & Salmivalli, C. (2009). Empirical test of bullies' status goals: Assessing direct goals, aggression, and prestige. *Aggressive Behavior*, *35*, 57–67.

Smith, D. A., & Brame, R. (2003). Tobit models in social science research. Some limitations and a more general alternative. *Sociological Methods & Research*, *31*, 364–388.

Tobin, J. (1953). Estimation for relationships with limited dependent variables. *Econometrica*, *26*, 24–36.

Veenstra, R., Lindenberg, S., Zijlstra, B. J. H., De Winter, A. F., Verhulst, F. C., & Ormel, J. (2007). The dyadic nature of bullying and victimization: Testing a dual perspective theory. *Child Development*, *78*, 1843–1854.

Child Development, March/April 2010, Volume 81, Number 2, Pages 487–502

Parent Discipline Practices in an International Sample: Associations With Child Behaviors and Moderation by Perceived Normativeness

Elizabeth T. Gershoff
University of Texas at Austin

Andrew Grogan-Kaylor
University of Michigan

Jennifer E. Lansford
Duke University

Lei Chang
Chinese University of Hong Kong

Arnaldo Zelli
Istituto Universitario di Scienze Motorie

Kirby Deater-Deckard
Virginia Polytechnic Institute and State University

Kenneth A. Dodge
Duke University

This study examined the associations of 11 discipline techniques with children's aggressive and anxious behaviors in an international sample of mothers and children from 6 countries and determined whether any significant associations were moderated by mothers' and children's perceived normativeness of the techniques. Participants included 292 mothers and their 8- to 12-year-old children living in China, India, Italy, Kenya, Philippines, and Thailand. Parallel multilevel and fixed effects models revealed that mothers' use of corporal punishment, expressing disappointment, and yelling were significantly related to more child aggression symptoms, whereas giving a time-out, using corporal punishment, expressing disappointment, and shaming were significantly related to greater child anxiety symptoms. Some moderation of these associations was found for children's perceptions of normativeness.

The study of parents' disciplinary practices has been central to developmental psychology for at least the last half century, tracing back to Sears, Maccoby, and Levin's (1957) landmark book, *Patterns of Child Rearing*. Their study of several hundred White working- and middle-class parents in and around Boston in the 1950s provided the first detailed data on what discipline techniques parents use, why they use them, and how discipline is linked with children's behaviors. Theirs was one of the first studies to document that, contrary to predictions that controlling parents would have better behaved children, in fact the opposite was true. In the ensuing 50 years, researchers have taken these same themes and tested relations among them in increasingly sophisticated models. What the field has not done consistently is to expand the populations of interest beyond families in the United States, or, in many cases, beyond White, middle-class parents.

The overreliance on children and parents from North America in research on parenting has led to the criticism that the theories and recommendations developed from North American samples are not universal (as they are often assumed to be) and may not apply to parents and children in other

---

Data collection for this work was funded by the Josiah Charles Trent Memorial Foundation and the Duke University Center for Child and Family Policy. We are grateful to the children and parents who participated in this research and to Dario Bacchini, Anna Silvia Bombi, Nandita Chaudhary, Paul Oburu, Kerstin Palmérus, Concetta Pastorelli, Lina Ramos, and Sombat Tapanya for their contributions to data collection. J.E.L. acknowledges support of NICHD Grant R01HD054805. K.A.D. acknowledges support of NIDA Grants K05DA015226 and P30DA023026.

Correspondence concerning this article should be addressed to Elizabeth T. Gershoff, University of Texas at Austin, Department of Human Development and Family Sciences, 1 University Station A2700, Austin, TX 78712. Electronic mail may be sent to liz.gershoff@austin.utexas.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8102-0007

countries and cultures (Dasen & Mishra, 2000; Rogoff, 2003). Cultures differ in the value they place on different child qualities and behaviors and in their beliefs about which parenting practices will promote these qualities and behaviors (Bornstein, 1995; Mistry, Chaudhuri, & Diez, 2003; Ripoll-Nuñez & Rohner, 2006; Super & Harkness, 1986). Cultures also differ in their evaluations of what parenting practices are considered to be either physically or emotionally abusive (Korbin, 2003).

Links between cultural values and parents' preferred childrearing practices have indeed been documented. A study of mothers in China and India found that differences in the two groups' use of authoritative and authoritarian parenting practices were related to relative differences in their valuing of goals of social emotional development and honor of the family (Rao, McHale, & Pearson, 2003). An analysis of ethnographies from 186 preindustrial societies found corporal punishment to be most likely among societies high in social stratification, political integration, long-term use of an alien currency, undemocratic political decision making, and a culture of violence (Ember & Ember, 2005). Parents in countries with more collectivist orientations, such as China, tend to emphasize the importance of behavioral inhibition and subjugation of individual wants for the good of the family, whereas parents in Western individualist cultures typically value assertiveness and independence in their children (Rothbaum, Morelli, Pott, & Liu-Constant, 2000).

Much of the research on discipline to date, both domestically and internationally, has focused on parents' use of corporal punishment, most likely because corporal punishment is a controversial form of discipline (Gershoff, 2002). However, corporal punishment is but one of many disciplinary techniques parents use in trying to control their children's negative behavior and promote their positive behavior. Although there is great variation in usage, with some parents using it everyday and some never using it, American parents tend to report using corporal punishment relatively rarely, peaking at one and a half times per month when children are 2 years old and decreasing to less than once per month by the time children are 12 years old (Straus & Stewart, 1999). In a recent longitudinal study of parents' use of 10 different disciplinary tactics, corporal punishment was one of the three least common techniques used by parents of preschoolers (Socolar, Savage, & Evans, 2007). Most common were monitoring child behavior, talking with children, distracting, and modeling.

In order to understand parental discipline we must expand our focus to the full range of parents' disciplinary techniques, particularly those they use more frequently than corporal punishment. There is very little research on the broad array of parental discipline techniques that are used in non-U.S. countries. The theories behind whether various discipline techniques should or should not be effective in promoting appropriate child behavior have largely been developed by researchers in North America (e.g., Bornstein, 1995; Dodge & Pettit, 2003; Grusec & Goodnow, 1994; Hoffman, 1983; Lepper, 1983). Whether parents in other countries differ in their preferences for disciplinary techniques and whether these techniques have similar associations with child behavior across countries remain open questions.

One reason discipline techniques may not have similar effects on children across countries is that the effects may depend on the extent to which the techniques are normative within a culture (Deater-Deckard & Dodge, 1997). When children perceive a discipline technique to be normative within their culture or community, they may be less likely to evaluate their parents' use of it as aberrant or objectionable. Discipline is most likely to be effective when children perceive it to be fair and reasonable (Grusec & Goodnow, 1994); if children believe the discipline they received to be out of the range of what their peers receive, they will reject and thus fail to internalize the parents' intended disciplinary message. Discipline that is perceived by children to be nonnormative is likely to be ineffective in promoting appropriate child behavior and may inadvertently lead to increases in problematic child behavior.

Parents' perceptions regarding the normativeness of the techniques they use may also moderate the effects of the techniques on children. When parents perceive a technique to be normative, they may feel more confident and justified in using it and thus may be more likely to use it in a planful and controlled, rather than impulsive and unregulated, manner. The planful, instrumental use of a disciplinary technique is thought to be less likely to cause fear or anxiety in children or to evoke reactive aggression from children (Holden, Miller, & Harris, 1999; Straus & Mouradian, 1998).

The Present Study

In order to advance our understanding of discipline in countries beyond the United States, we collaborated with colleagues in six understudied countries, namely, China, India, Italy, Kenya, the Philippines, and Thailand. This sample of countries was diverse on several demographic dimensions, including predominant race/ethnicity and predominant religion. They also varied on several key social and economic indicators, including: gross national income (from highest to lowest: Italy, Thailand, China, Philippines, India, and Kenya, with Italy's income 49 times as large as that of Kenya; World Bank, 2008), individualist versus collectivist social orientation (from most to least individualistic: Italy, India, Philippines, Kenya, China, and Thailand; Hofstede, 2001), and overall child well-being, indexed as the mortality rate for children < 5 years old, the underweight rate for children < 5 years old, and the percent of primary-school-aged children not enrolled in school (ranked from best to worst: Italy, China, Thailand, Philippines, Kenya, and India; Save the Children, 2008). Although Italy is first in all three of these rankings, the order of the remaining five countries is not the same for all three indicators, indicating that these countries vary from one another on multiple dimensions. By studying families from each of these countries, we created a diverse international sample with which to examine parent discipline and child behavior.

In a previous study with these data (Lansford et al., 2005), we examined the extent to which the normativeness of harsh corporal punishment (which included spanking, slapping, grabbing, shaking, and beating up) in each of the country subsamples moderated the association between frequency of harsh corporal punishment and children's aggression and anxiety symptoms. More frequent use of harsh corporal punishment was associated with more child aggression and anxiety across all six countries, but the strength of these associations was strongest in country subsamples in which harsh corporal punishment was least normative (Lansford et al., 2005).

Whether children's and mothers' perceptions of cultural normativeness play similar moderational roles for other disciplinary techniques remains unknown. Thus, for this study, we used the same international sample to address two new research questions, namely: (a) When multiple discipline techniques are considered simultaneously, which forms of discipline emerge as having the strongest associations with children's aggressive and anxious behaviors? and (b) Are any significant associations between discipline practices and child behaviors moderated by the extent to which mothers and children perceive these practices to be normative within their communities? We hypothesized that high normativeness will only moderate techniques that are most likely to be experienced as negative events by the child, namely, corporal punishment and yelling. We expected that more benign techniques, such as reasoning, will not be affected by normativeness, although we left open the possibility that less aversive forms of discipline may actually have a stronger impact on children if they are less normative and thus experienced by children as salient positive events.

Method

Participants

The sample of 292 families for this study was drawn from a complete project sample of 336 children and their mothers from seven metropolitan areas in six countries: China (Beijing), India (Delhi), Italy (Rome and Naples), Kenya (Rachuonyo District of Nyanza province), Philippines (Manila), and Thailand (Chiang Mai). Within each country, samples were recruited through public schools in neighborhoods that were identified by local contacts as primarily middle class. After permission was obtained from appropriate local authorities (e.g., principals, teachers), children and their parents were invited to participate in a study of how parents discipline children and how parents' behavior affects how children learn and grow. For a family to be eligible for inclusion in the study, both the mother and the child had to be willing to complete the interview.

Although the target age group for the study was mothers and children aged 8–12 years old, the range of ages of children who actually participated was 6–17 ($M = 10.57$, $SD = 1.86$). Because discipline techniques are used differently with children in early childhood, middle childhood, and adolescence and may have different effects depending on the developmental stage in which they are used (Collins, Madsen, & Susman-Stillman, 2002; Giles-Sims, Straus, & Sugarman, 1995), we restricted our sample for these analyses to the target age group of middle childhood (range = 8–12; $M = 10.22$, $SD = 1.30$). This restriction resulted in our sample of 292 children and their mothers (87% of the original sample) for the analyses reported here. The number of participants and percent of the children who were female in each country-level sample were: China: $n = 49$, 47% girls; India: $n = 36$, 58% girls; Italy: $n = 73$, 43% girls; Kenya: $n = 17$, 53% girls; Philippines: $n = 50$, 38% girls; and Thailand: $n = 58$, 57% girls.

Mothers ranged in age from 20 to 59 years (M = 38.66, SD = 5.87). Although there are ethnic minorities in these countries, the participants did not identify themselves as being members of any ethnic minority groups. In 95% of the cases, the biological mother was interviewed; in the remaining cases, a female adoptive parent or relative who was the child's primary caregiver was interviewed. Henceforth, we will refer to the parent respondents as "mothers." Within each country, the samples were considered primarily middle class and had similar standings in terms of within-country socioeconomic status. Mothers reported their families' annual income in local currency using ranges that reflected income distributions within a particular country. The median annual income range (converted to U.S. dollars) in each country subsample was as follows: (a) Chinese subsample median = \$2,172–\$5,796 (which may be an underestimate because it does not include bonuses that many Chinese employees earnin addition to their base salary), (b) Indian subsample median = \$13,728–\$16,464, (c) Italian subsample median = \$32,585–\$37,799, (d) Kenyan subsample median = \$1,560–\$2,352, (e) Philippines subsample median = \$3,306–\$4,404, and (f) Thai subsample median = \$3,036–\$15,180. The differences in the width of these income ranges are reflective of the, sometimes widely varying, range of incomes in the communities from which our sub-samples were drawn. Differences in relative income between countries were accounted for by our use of fixed effects models.

Procedure

All interviews were conducted in the homes of the participants by trained interviewers who were natives of the country. Most of the interviewers were graduate students in psychology and all conducted pilot trials before the actual interview. Before the formal interview began, both the mother and the child gave consent to participate in the interview. During each interview, which lasted about 1 hr, the interviewer read aloud each question individually from the printed questionnaire. Because not all participants were literate, all questions were asked orally. After each question was read, the participant verbalized or pointed to the answer on the printed questionnaire or on a separate pictorial response card and responses were recorded by the interviewer. When any participant expressed confusion about a given question, the interviewer provided only brief explanations or elaborations on it, typically consisting of paraphrasing the question only. The mother and the child were interviewed separately in different rooms or at different times so that they could not hear each other's responses. To minimize social desirability, during the course of the interviews participants were reminded that their answers would be kept confidential, that there were no right or wrong answers to the questions, and that it was important for them to answer as honestly as possible. Country-level differences in social desirability were accounted for, along with all other unmeasured differences between countries, by our use of fixed effects models in the analyses described below.

Measures

A procedure of translation and back-translation was used to ensure the linguistic and conceptual equivalence of measures across languages. The translators were fluent in English and the target language. Translators were asked to note places in the research instruments that did not translate well, were inappropriate for the age groups in the study, or were culturally insensitive. Any problems noted were resolved through discussions among the translators and investigators. Additionally, the appropriateness of the terms used to assess each form of discipline (e.g., *spank*, *slap*, or *hit*) was evaluated through qualitative interviews with a separate pilot sample of mothers in each country. English versions of the measures were administered in India and the Philippines, countries where English is an official language. Measures in the other countries were administered in Mandarin Chinese, Italian, Dholuo (Kenya), and Thai.

Maternal discipline techniques. The frequency with which mothers reported using each of 11 different types of discipline was assessed through a measure that was developed for the present study, adapted from other instruments that assess parents' discipline strategies (Deater-Deckard, Dodge, & Sorbring, 2005; Straus, 1979). The discipline questionnaires were introduced by saying, "All children misbehave sometimes. One of the things we need to learn about is what parents do to deal with their children's misbehavior. Think about when your child has misbehaved in the last year." Mothers were then asked a series of questions specific to each discipline technique, such as "How frequently do you teach your child about good and bad behavior? Like, it's not nice to hit, or it's polite to say 'thank you'." Mothers reported how frequently they used each discipline technique on a 5-point scale (1 = *never*, 2 = *less than once a month*, 3 = *about once a month*, 4 = *about once a week*, 5 = *almost every day*). The 11 discipline techniques as they were described in the interviews, listed in the order they were presented to the mothers and followed by the abbreviations we will use hereafter, were: (1) "teach your child about good and bad behavior" (*teach about good and bad behavior*); (2) "get your child to apologize or make amends" (*get child to apologize*); (3) give child a time-out or send him/-her to his/her room" (*give a time-out*); (4) "take away privileges" (*take away privileges*); (5) "spank, slap, or hit your child" (*use corporal punishment*); (6) "say you are disappointed with your child, or say that his/her misbehavior hurt your feelings" (*express disappointment*); (7) "tell your child s/he should be ashamed of her/himself" (*shame*); (8) "raise your voice, yell, or scold your child" (*yell or scold*); (9) "tell your child you won't love him/her if he/she acts that way again" (*withdraw love for misbehavior*); (10) "threaten your child with some punishment" (*threaten punishment*); and (11) "promise a treat or privilege to your child for good behavior" (*promise a treat or privilege*). Pooled sample means and country means are reported in Table 1.

Perceived normativeness of discipline techniques. Both mothers and their children were asked a question about how normative they perceived each discipline technique to be in their community. For each of the 11 techniques listed above, mothers were asked a question beginning with the stem, "How frequently do other parents in your community . . . " Similarly, children were asked a question beginning with the stem, "When kids do something wrong how often do their parents . . . " for each discipline technique. Both parents and children reported how often they thought other parents in their community used each technique according to the same scale mothers used to report the frequency with which they used each discipline technique, namely: 1 = *never*, 2 = *less than once a month*, 3 = *about once a month*, 4 = *about once a week*, 5 = *almost every day*. Mothers' and children's ratings of normativeness were significantly correlated across 10 of the 11 discipline techniques, with the significant correlations ranging from $r(273) = .13$, $p < .05$, for *teach about good and bad behavior*, to $r(274) = .46$, $p < .001$, for *give a time-out*. Mothers' and children's ratings of the extent to which other parents *get children to apologize* were the only ratings not significantly correlated, $r(275) = .02$, $p = .71$. Pooled sample means and country means for mother and child ratings of the normativeness of each discipline technique are reported in Table 1.

Child behaviors. Children's aggression and anxiety behaviors were obtained from mothers' reports using the aggression subscale (20 items) and the Anxiety/Depression subscale (14 items) of the Child Behavior Checklist (CBCL; Achenbach, 1991a). Self-reported symptoms were obtained through youth responses to the Aggression subscale (19 items) and the Anxiety/Depression subscale (16 items) of the Youth Self-Report (YSR; Achenbach, 1991b). The CBCL and YSR have been translated into at least 69 languages, and over 5,000 published studies have used these measures with at least 60 cultural groups (Achenbach, 2004; Crijnen, Achenbach, & Verhulst, 1997). The Achenbach measures have been used previously in all of the countries involved in the present study: China (e.g., Dong, Wang, & Ollendick, 2002; Guo et al., 2000), India (Gill & Kang, 1995; Shenoy, Kapur, & Kaliaperumal, 1998), Italy (Artigas, 1999), Kenya (Weisz, Sigman, Weiss, & Mosk, 1993), the Philippines (Florencio, 1988), and Thailand (Weisz, Suwanlert, et al., 1993). In the current study, both subscales had strong internal reliability for the entire sample, for mother and child reports, respectively: aggression $\alpha s = .80$, $.80$; anxiety $\alpha s = .74$, $.75$. Within-country internal consistencies for mother and child reports of aggression were: China, $\alpha s = .79$, $.80$; India, $\alpha s = .58$, $.54$; Italy, $\alpha s = .67$; Kenya, $\alpha s = .73$, $.81$; the Philippines, $\alpha s = .84$, $.79$; Thailand, $\alpha s = .80$, $.83$. For the anxiety subscales, internal consistencies for the mother- and child-reports were: China, $\alpha s = .88$, $.76$; India, $\alpha s = .50$, $.71$; Italy, $\alpha s = .71$; Kenya, $\alpha s = .64$, $.81$; the Philippines, $\alpha s = .81$, $.82$; Thailand, $\alpha s = .69$, $.71$. Because the YSR was only added to the study after the Italian families had participated, we do not have youth self-ratings of aggression and anxiety for the Italian subsample.

Analyses

We utilized two complementary analytic approaches in order to account for between-country differences. For each dependent variable, we first estimated a multilevel model, which is suitable for estimating the association of a set of independent variables with a dependent variable of interest. Multilevel models are commonly employed in situations where data are clustered, as in this case where data are clustered by country (Hox, 2002; Raudenbush & Bryk, 2002). In the multilevel models used in this analysis, the clustering of data is accounted for by allowing the intercept term of the

Table 1
Within-Country Means of Reported Use of and Perceived Normativeness of 11 Discipline Techniques

		Pooled sample mean	Country					
			China	India	Italy	Kenya	Philippines	Thailand
Teach about good and bad behavior	Frequency of use	4.42 (0.88)	4.29 (0.96)	4.60 (0.54)	4.36 (0.89)	3.53 (1.18)	4.76 (0.62)	4.43 (0.90)
	Normativeness (M)	3.77 (1.26)	3.84 (1.36)	4.11 (0.96)	3.77 (1.26)	3.00 (1.06)	4.36 (0.88)	3.10 (1.51)
	Normativeness (C)	4.20 (1.18)	3.79 (1.25)	4.39 (0.81)	3.79 (1.43)	3.94 (1.25)	4.67 (0.92)	4.57 (0.90)
Get child to apologize	Frequency of use	3.50 (1.32)	3.10 (1.34)	3.04 (1.09)	3.31 (1.48)	2.65 (1.46)	4.18 (0.87)	4.07 (1.06)
	Normativeness (M)	3.10 (1.31)	2.96 (1.45)	2.91 (1.00)	3.06 (1.33)	2.65 (1.11)	3.76 (1.04)	2.94 (1.53)
	Normativeness (C)	3.43 (1.38)	2.69 (1.21)	3.27 (1.19)	3.35 (1.56)	3.00 (1.41)	4.39 (0.73)	3.60 (1.36)
Give a time-out	Frequency of use	1.94 (1.29)	1.12 (0.44)	1.82 (1.12)	1.53 (0.95)	1.82 (1.07)	3.86 (0.95)	1.64 (1.00)
	Normativeness (M)	2.18 (1.27)	1.35 (0.80)	2.20 (1.01)	2.20 (1.11)	2.00 (1.50)	3.80 (0.78)	1.35 (0.66)
	Normativeness (C)	2.12 (1.35)	1.57 (1.02)	1.84 (1.21)	2.29 (1.36)	1.59 (1.00)	3.80 (0.89)	1.31 (0.73)
Take away privileges	Frequency of use	2.35 (1.21)	2.02 (1.09)	2.13 (1.16)	2.33 (1.24)	1.47 (0.87)	3.20 (0.95)	2.33 (1.22)
	Normativeness (M)	2.47 (1.22)	2.02 (1.14)	2.38 (1.25)	2.93 (1.08)	1.65 (1.00)	3.24 (0.87)	1.77 (1.15)
	Normativeness (C)	2.36 (1.36)	1.94 (1.28)	2.00 (1.38)	2.71 (1.35)	1.65 (1.46)	3.28 (1.28)	2.02 (1.18)
Use corporal punishment	Frequency of use	2.03 (1.02)	1.86 (1.00)	2.31 (1.08)	2.15 (0.94)	3.06 (0.97)	2.04 (1.09)	1.52 (0.71)
	Normativeness (M)	2.47 (1.16)	1.66 (0.83)	2.68 (1.03)	2.97 (1.12)	3.41 (1.06)	2.76 (0.98)	1.71 (0.97)
	Normativeness (C)	2.29 (1.33)	1.75 (1.16)	2.38 (1.11)	3.21 (1.36)	3.82 (1.13)	1.86 (0.98)	1.48 (0.75)
Express disappointment	Frequency of use	2.51 (1.18)	2.10 (1.21)	2.87 (0.92)	2.56 (1.22)	2.59 (1.42)	2.44 (1.28)	2.57 (1.03)
	Normativeness (M)	2.59 (1.23)	1.89 (0.98)	3.05 (0.94)	2.88 (1.15)	3.00 (1.41)	3.02 (1.23)	1.84 (1.12)
	Normativeness (C)	2.39 (1.30)	1.96 (1.15)	2.69 (1.28)	2.63 (1.40)	2.76 (1.48)	2.24 (1.32)	2.24 (1.11)
Shame	Frequency of use	2.21 (1.17)	1.82 (1.20)	2.71 (1.06)	2.36 (1.26)	2.71 (1.05)	1.96 (1.01)	2.02 (1.08)
	Normativeness (M)	2.38 (1.21)	1.72 (1.03)	2.95 (1.01)	2.67 (1.27)	2.88 (0.86)	2.69 (1.23)	1.60 (0.90)
	Normativeness (C)	2.41 (1.33)	1.71 (1.05)	2.56 (1.12)	3.01 (1.46)	3.75 (1.13)	2.00 (1.10)	2.12 (1.17)
Yell or scold	Frequency of use	3.26 (1.27)	2.82 (1.18)	3.29 (0.92)	4.21 (1.05)	3.06 (1.34)	3.06 (1.42)	2.69 (1.06)
	Normativeness (M)	3.35 (1.26)	2.45 (1.23)	3.54 (0.70)	4.27 (0.96)	3.06 (1.03)	3.34 (1.17)	2.78 (1.35)
	Normativeness (C)	2.86 (1.37)	2.02 (1.16)	2.96 (1.21)	3.59 (1.31)	3.88 (1.45)	2.94 (1.20)	2.28 (1.20)
Withdraw love for misbehavior	Frequency of use	1.83 (1.15)	2.06 (1.36)	2.31 (1.24)	1.41 (0.79)	2.06 (1.43)	1.50 (0.81)	1.98 (1.19)
	Normativeness (M)	2.00 (1.10)	1.89 (1.09)	2.54 (1.09)	1.87 (1.08)	2.24 (1.25)	2.02 (0.92)	1.67 (1.09)
	Normativeness (C)	1.71 (1.08)	1.41 (0.93)	1.91 (1.04)	1.67 (1.05)	2.24 (1.25)	1.80 (1.17)	1.65 (1.08)
Threaten punishment	Frequency of use	2.64 (1.28)	1.94 (1.20)	2.38 (1.07)	2.85 (1.35)	4.12 (1.22)	2.82 (1.12)	2.59 (1.12)
	Normativeness (M)	2.78 (1.36)	1.63 (0.95)	2.61 (1.04)	3.19 (1.32)	4.53 (1.01)	3.20 (1.05)	2.38 (1.34)
	Normativeness (C)	2.32 (1.39)	1.57 (1.04)	1.91 (1.10)	2.53 (1.42)	4.71 (0.77)	2.79 (1.15)	1.93 (1.17)
Promise a treat or privilege	Frequency of use	3.07 (1.33)	2.46 (1.34)	3.56 (0.97)	2.65 (1.40)	2.12 (1.62)	3.80 (0.88)	3.34 (1.16)
	Normativeness (M)	3.05 (1.27)	2.42 (1.30)	3.72 (0.83)	3.40 (1.16)	2.31 (1.45)	3.36 (0.88)	2.48 (1.43)
	Normativeness (C)	3.28 (1.42)	1.92 (1.15)	3.49 (0.85)	3.60 (1.46)	1.71 (1.36)	3.94 (1.00)	3.78 (1.17)

Note. Scale values are: 1 = *never,* 2 = *less than once a month,* 3 = *about once a month,* 4 = *about once a week,* 5 = *almost every day.* M = mother report; C = child report. Means are presented with standard deviations in parentheses.

model to randomly vary across countries (Raudenbush & Bryk, 2002; Stata Corporation, 2005). Multilevel models estimate both within country and between country differences (Raudenbush & Bryk, 2002). In other words, there may be differences in parenting and child behavior between two families in the same country, but there may also be differences in parenting and child behavior across countries. Both sources of variation are incorporated into the parameters of the multilevel models.

However, one drawback of using multilevel models for this project is that with a relatively small number of Level 2 units (six countries), the distributional assumptions necessary for estimation of a random intercept may not be satisfied (Maas & Hox, 2004). Therefore, for each dependent variable we also estimated a fixed effects model. As the name of these models suggests, country-specific intercepts are estimated as fixed, rather than random, and thus the models are not subject to the same distributional concerns as multilevel models. In addition, fixed effect models are able to control for both observed and unobserved country subsample-level differences (Allison, 2005). One potential limitation of fixed effects models is that all between-country differences are subsumed into the country-specific terms, and the parameter estimates of the model only reflect differences between individuals in the same countries. In the present article, we conducted parallel multilevel and fixed effects analyses in order to identify and understand potential moderation by country of residence in the most robust manner.

Because the country subsamples were small and were not chosen to be representative of their respective countries, we cannot extrapolate any subsample differences up to country-level differences. Our emphasis in these analyses is on whether country of residence explains any significant portion of the associations of 11 discipline techniques with child mental health and behavior. We cannot make conclusions about any given country included in our study, but rather treat ours as an international sample with which to determine whether the associations of parental discipline with child mental health and behavior are culture- (i.e., country-) specific or potentially universal.

Results

Discipline Techniques in Relation to Children's Aggression and Anxiety Across Countries

Table 2 summarizes the results for the models in which mother- and child-reported aggression were regressed on the 11 discipline techniques considered simultaneously; all models control for children's gender and age. The significant intercepts for country variance in both multilevel models (intercepts = 0.48 and 0.46, both *p*s < .01) indicate that there are country-level differences in average levels of child aggression across the countries,

Table 2
Results of Multilevel and Fixed Effects Models Predicting Children's Aggression Symptoms From Mothers' Reported Frequency of Using 11 Discipline Techniques in Six Countries

	Child aggression—Mother reports				Child aggression—Self reports			
	Multilevel model		Fixed effects model		Multilevel model		Fixed effects model	
	β	*SE*	β	*SE*	β	*SE*	β	*SE*
Fixed effects								
Intercept	−1.72**	0.52	−1.57**	0.49	−0.54	0.07	−0.31	0.70
Child gender[a]	−0.25**	0.10	−0.24**	0.10	−0.03	0.13	−0.02	0.13
Child's age	−0.00	0.04	0.00	0.04	−0.03	0.05	−0.03	0.05
Teach about good and bad behavior	0.02	0.06	0.01	0.06	−0.12	0.09	−0.13	0.09
Get child to apologize	−0.01	0.05	−0.02	0.05	−0.04	0.07	−0.05	0.07
Give a time-out	0.05	0.05	0.04	0.05	0.10	0.07	0.08	0.07
Take away privileges	0.06	0.05	0.05	0.05	0.02	0.07	−0.01	0.07
Use corporal punishment	0.17**	0.05	0.18**	0.05	0.15	0.08	0.16*	0.08
Express disappointment	0.13**	0.05	0.13**	0.05	0.09	0.07	0.10	0.07
Shame	0.08	0.05	0.09	0.05	−0.02	0.07	−0.00	0.08
Yell or scold	0.14**	0.05	0.13**	0.05	0.21**	0.06	0.20**	0.06
Withdraw love for misbehavior	0.04	0.05	0.04	0.05	0.04	0.07	0.04	0.07
Threaten punishment	0.05	0.05	0.06	0.05	0.01	0.07	0.04	0.07
Promise a treat or privilege	−0.02	0.04	−0.02	0.04	−0.00	0.06	−0.02	0.06
Random effects for multilevel models								
Intercept for country variance	0.48**	0.17	—		0.46**	0.22	—	
Residual variance	0.76**	0.03	—		0.90**	0.05	—	
Fit indices for fixed effects models								
R^2	—		0.47		—		0.28	
Root mean square error	—		0.76		—		0.90	
F			9.33*** (13, 256)				3.47*** (13, 188)	

Note. Standardized coefficients and their standard errors are presented. The models for mother-rated outcomes included 275 families in the six countries. Because the Youth Self-Report was not asked of children in the Italian sample, the models for child-rated outcomes include 206 families and five countries.

[a] 0 = boys, 1 = girls.

**p* < .05. ***p* < .01. ****p* < .001.

while the significant R^2s from the fixed effects models (R^2s = 0.47 and 0.28, both ps < .001) indicate that parent discipline techniques do a good job explaining the variance in child aggression, with 47% of the variance in mother-reported aggression explained and 28% of child-reported aggression explained. Mothers rated their boys as more aggressive than did mothers of girls; however, there were no gender differences in child-reported symptoms of aggression.

The results from the multilevel and fixed effects models were consistent—in all but one case (child-reported aggression as a function of mothers' *use corporal punishment* in the multilevel model), all coefficients significant in the multilevel model were also significant in the fixed effects model. Across both the multilevel and fixed effects models, the more often mothers *use corporal punishment*, the more aggressive they reported their children were (βs = .17 and .18, $p < .01$, for the multilevel and fixed effects models, respectively). The frequency with which mothers yell or scold was significantly associated with mothers' ratings of children's aggression (βs = .14 and .13, $p < .01$, in the multilevel and fixed effects models). Both *use corporal punishment* and *yell or scold* similarly were associated with children's self-reported aggression, although corporal punishment was not significant in the multilevel model (β = .15, *ns*, and β = .16, $p < .05$, for *use corporal punishment* in the multilevel and fixed effects models; βs = .21 and .20, $p < .01$, for *yell or scold*). The extent to which mothers reported that they *express disappointment* was associated with more mother-reported aggression (β = .13, $p < .01$, in both models), but not with more child-reported aggression (βs = .09 and .10, *ns*). The remaining eight discipline techniques were not significantly associated with child aggression in any of the models.

Results from the multilevel and fixed effects models for the dependent variables of mother- and child-reported anxiety symptoms while controlling for child gender and age are presented in Table 3. As with the models for child aggression symptoms, the results from the multilevel and fixed effects models for child anxiety symptoms were generally the same. The one difference was again for *use corporal punishment*, which was significantly associated with child-reported anxiety symptoms in the fixed effects model (β = .17, $p < .05$) but not in the multilevel model (β = .15, *ns*). The 11 discipline techniques did a fair job accounting for child anxiety symptoms, explaining 34% of the variance in mother-reported symptoms and 19% of the variance in child-reported symptoms. Different discipline techniques were significant in relation to mother- versus child-reported anxiety symptoms. In the mother-reported models, the frequency with which mothers *give a time-out* was significantly correlated with child anxiety symptoms (βs = .13 and .14, $p < .05$, respectively, for the multilevel and fixed effects models) as was the frequency with which mothers *shame* their children (β = .12, $p < .05$, in both models). In the models with children's own reports of their anxiety symptoms as the dependent variable, the frequency with which mothers report they *use corporal punishment* was significant in the fixed effects model only (β = .17, $p < .05$) while *express disappointment* was significantly associated with children's anxious behaviors in both models (β = .15, $p < .05$, in both models). Seven discipline techniques were not associated with child anxiety symptoms in any of the four models tested.

For all forms of discipline in the multilevel models statistically significant at $p < .05$, models were re-estimated with random slope parameters in order to see whether there was variation in the significant slopes by country. In only two cases did a likelihood ratio test indicate that there was statistically significant variation by country in the slope parameter for a specific discipline type. There was statistically significant variation in the slope for the association of *use corporal punishment* with child-rated aggression and in the slope for the association of *express disappointment* with mother-rated aggression. Both of these models were then re-estimated with country differences treated as fixed in order to examine country-specific slopes.

A fixed effects model examining country-specific slopes for the association of *use corporal punishment* with child-rated aggression did not find any statistically significant country-level variation in the slope. This finding indicates that the variation in slope found in the multilevel model might have been due to the distributional assumptions required by multilevel models or to unobserved differences between countries. In contrast, results from these fixed effects models suggested that *express disappointment* was more strongly associated with increased levels of mother-rated aggression among families in the Chinese, Italian, Philippine, and Thai subsample than among families in the Kenyan subsample. *Express disappointment* was more strongly associated with mother-rated aggression for families in the Chinese subsample than for families in the Indian subsample. However, across all country subsamples, more frequent use of

Table 3

Results of Multilevel and Fixed Effects Models Predicting Children's Anxiety Symptoms From Mothers' Reported Frequency of Using 11 Discipline Techniques in Six Countries

	Child anxiety—Mother report				Child anxiety—Self report			
	Multilevel model		Fixed effects model		Multilevel model		Fixed effects model	
	β	*SE*	β	*SE*	β	*SE*	β	*SE*
Fixed effects								
Intercept	−2.79**	0.56	−2.72**	0.55	−1.37	0.76	−1.25	0.74
Child gender[a]	0.18	0.11	0.17	0.11	0.26	0.14	0.26	0.14
Child's age	0.10*	0.04	0.11*	0.04	0.05	0.05	0.05	0.05
Teach about good and bad behavior	0.03	0.07	0.03	0.07	−0.05	0.10	−0.07	0.10
Get child to apologize	−0.03	0.05	−0.03	0.05	−0.02	0.08	−0.03	0.08
Give a time-out	0.13*	0.06	0.14*	0.06	0.04	0.07	0.04	0.08
Take away privileges	0.07	0.05	0.06	0.05	0.09	0.07	0.07	0.07
Use corporal punishment	0.11	0.06	0.12	0.06	0.15	0.08	0.17*	0.08
Express disappointment	0.08	0.06	0.08	0.06	0.15*	0.07	0.15*	0.07
Shame	0.12*	0.06	0.12*	0.06	−0.06	0.08	−0.06	0.08
Yell or scold	0.09	0.05	0.08	0.05	0.07	0.07	0.05	0.07
Withdraw love for misbehavior	0.09	0.06	0.10	0.06	−0.10	0.07	−0.11	0.07
Threaten punishment	−0.04	0.06	−0.03	0.06	0.08	0.07	0.11	0.08
Promise a treat or privilege	0.01	0.05	0.02	0.05	−0.03	0.07	−0.04	0.07
Random effects for multilevel models								
Intercept for country variance	0.41**	0.15	—		0.38**	0.18		
Residual variance	0.85**	0.04	—		0.96**	0.05		
Fit indices for fixed effects models								
R^2	—		0.34		—		0.19	
Root mean square error	—		0.85		—		0.96	
F			5.76*** (13, 256)				2.05* (13, 188)	

Note. Standardized coefficients and their standard errors are presented. The models for mother-rated outcomes included 275 families in the six countries. Because the Youth Self-Report was not asked of children in the Italian sample, the models for child-rated outcomes include 206 families and five countries.

[a] 0 = boys, 1 = girls.

*p < .05. **p < .01. ***p < .001.

express disappointment was associated with higher levels of aggression.

Perceived Normativeness as a Moderator of Associations Among Discipline Techniques and Children's Aggression and Anxiety

To examine whether mothers' and children's perceptions of the normativeness of discipline techniques moderated their associations with child behavior, we focused on the discipline techniques that were significant main effects in either the multilevel or the fixed effects models for the child aggression and anxiety dependent variables in Tables 2 and 3. We thus examined the potential for perceived normativeness to moderate the statistically significant relations between 5 of the 11 discipline techniques and one or both of the child behaviors. Specifically, we focused on the associations of *give a time-out* with mother-rated anxiety, *use corporal punishment* with mother- and child-reported aggression and child-reported anxiety, *express disappointment* with mother-reported aggression and child-reported anxiety, *shame* with mother-rated anxiety, and *yell or scold* with mother- and child-reported aggression. For each of these nine discipline–outcome combinations, we ran the models separately for moderation by mother-perceived normativeness and by child-perceived normativeness. Employing a similar approach to that used above, each model was run as a multilevel model and then again as a fixed effects model; in all cases, the results were nearly identical for each approach. For ease of interpretation, we present and discuss only the results from the multilevel models.

Table 4 displays the results from the tests of moderation by mother-reported normativeness. Each of these models controlled for child age and child gender. In only one of the models was the main effect for mother-reported normativeness significant as well as the interaction term. The more normative mothers thought *express disappointment* was in their community, the more anxiety symptoms their children self-reported. The significant negative interaction term between the frequency with which mothers *express disappointment* and how normative they perceived it to be, β = −.12, *p* < .001, is graphed in Figure 1. The slope of the regression line between frequency with which mothers *express disappointment* and the number of anxiety symptoms reported by their children is steeper under conditions of low normativeness than under high normativeness. The intercept was higher for the high normativeness condition.

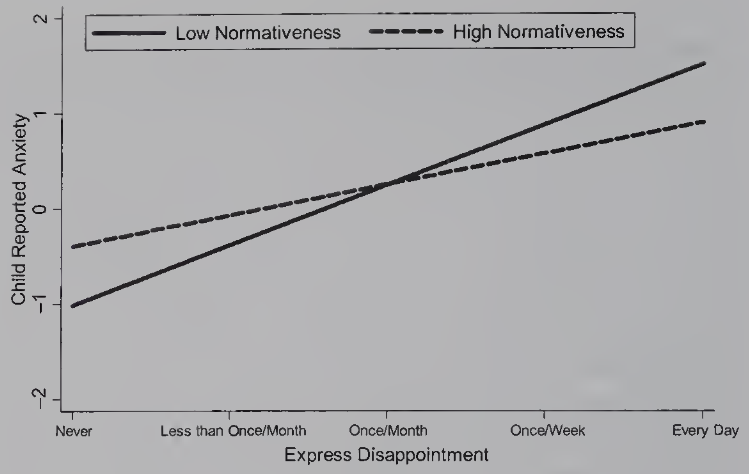


Figure 1. Plot of the link between frequency of mothers' expressions of disappointment and children's anxiety symptoms as moderated by mothers' perceptions of the normativeness of expressing disappointment.
Note. "Low normativeness" and "high normativeness" reflect the frequency with which mothers express disappointment at −1 and +1 *SD*, respectively, from the mean of mothers' reports of normativeness across the sample.

Children's perceived normativeness of various discipline techniques was a stronger moderator of the associations between discipline and child behavior; it was significant in five of the nine models tested (see Table 5). Children's perceptions of the normativeness of corporal punishment significantly moderated the associations of mothers' use of corporal punishment with children's aggression and anxiety, βs = −.07, −.14, and −.10. Graphs

Table 4
Results of Multilevel Models Predicting Children's Aggression and Anxiety Symptoms From Significant Discipline Techniques Moderated by Mothers' Perceived Normativeness

	Child aggression		Child anxiety	
	Mother report	Child report	Mother report	Child report
Give a time-out				
Frequency with which mother gives a time-out			.20 (.13)	
Mother's perceived normativeness of time-out			.00 (.10)	
Interaction			.01 (.04)	
Use corporal punishment				
Frequency with which mother uses corporal punishment	.28** (.12)	.29* (.15)		.19 (.16)
Mother's perceived normativeness of corporal punishment	.14 (.10)	−.06 (.13)		−.02 (.14)
Interaction	−.01 (.04)	.01 (.05)		.01 (.05)
Express disappointment				
Frequency with which mother expresses disappointment	.17 (.11)			.48*** (.14)
Mother's perceived normativeness of expressing disappointment	−.10 (.10)			.38** (.15)
Interaction	.03 (.03)			−.12*** (.05)
Shame				
Frequency with which mother shames			.06 (.11)	
Mother's perceived normativeness of shaming			−.08 (.11)	
Interaction			.05 (.04)	
Yell or scold				
Frequency with which mother yells or scolds	.21* (.12)	.17 (.15)		
Mother's perceived normativeness of yelling or scolding	−.02 (.12)	−.13 (.15)		
Interaction	.01 (.03)	.04 (.04)		

Note. All models controlled for child age and child gender. Values presented are standardized with standard errors in parentheses.
**p* < .05. ***p* < .01. ****p* < .001.

Table 5
Results of Multilevel Models Predicting Child Aggression and Anxiety Symptoms From Significant Discipline Techniques Moderated by Children's Perceived Normativeness

	Child aggression		Child anxiety	
	Mother report	Child report	Mother report	Child report
Give at time-out				
Frequency with which mother gives a time-out			.16 (.10)	
Child's perceived normativeness of time-out			.02 (.09)	
Interaction			.02 (.03)	
Use corporal punishment				
Frequency with which mother uses corporal punishment	.51*** (.11)	.49*** (.14)		.40*** (.15)
Child's perceived normativeness of corporal punishment	.17* (.09)	.57*** (.13)		.35** (.14)
Interaction	−.07* (.04)	−.14*** (.05)		−.10* (.05)
Express disappointment				
Frequency with which mother expresses disappointment	.35*** (.09)			.54*** (.11)
Child's perceived normativeness of expressing disappointment	.13 (.09)			.72*** (.12)
Interaction	−.03 (.03)			−.19*** (.04)
Shame				
Frequency with which mother shames			.24** (.10)	
Child's perceived normativeness of shaming			.01 (.10)	
Interaction			.00 (.03)	
Yell or scold				
Frequency with which mother yells or scolds	.44*** (.09)	.15 (.12)		
Child's perceived normativeness of yelling or scolding	.23** (.10)	.11 (.13)		
Interaction	−.06** (.03)	.02 (.04)		

Note. All models controlled for child age and child gender. Values presented are standardized with standard errors in parentheses.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

of these significant interactions are presented in Figure 2. The slope of the regression line under the condition of low normativeness was steeper than that for high normativeness in each case. These findings suggest that although more frequent corporal punishment is associated with more child- and mother-rated aggression and child-rated anxiety in both conditions, mothers' use of corporal punishment is less strongly associated with child behaviors when children believe it to be normative in their communities.

Similar results were found for moderation of *express disappointment* and *yell or scold*. In both cases, the interaction of child-perceived normativeness with the frequency with which mothers used each technique was negative ($\beta = -.19$, $p < .001$, and $\beta = -.06$, $p < .01$), indicating a decreased slope. The graphs of these interactions are displayed in the right half of Figure 2. As with the other interaction terms, the slopes of the regression lines under high-normativeness conditions are less steep than the slopes of the lines under the low-normativeness condition. Also of interest is that, similar to the other graphs, the intercept is higher under the high-normativeness condition.

Discussion

This study examined the disciplinary practices of mothers in six countries as well as the associations of these discipline techniques with child aggression and anxiety. In parallel analyses with country of residence treated as a random intercept (in multilevel models) and as a fixed parameter (in fixed effects models), we tested the extent to which the associations of maternal discipline techniques with children's aggressive and anxious symptoms are culture (country) specific or universal across countries. We also examined the extent to which perceived normativeness of a given discipline technique might moderate these links. Our goal in this manuscript was to examine links between a range of discipline techniques and child aggression and anxiety in a unique, international sample, with country differences treated by our choice of analytic techniques. In other words, our intention was not to treat this as an opportunity to make generalizations about parents in particular countries.

Our first research question was, once you take into account that parents differ by country in their preferred discipline techniques, is each form of

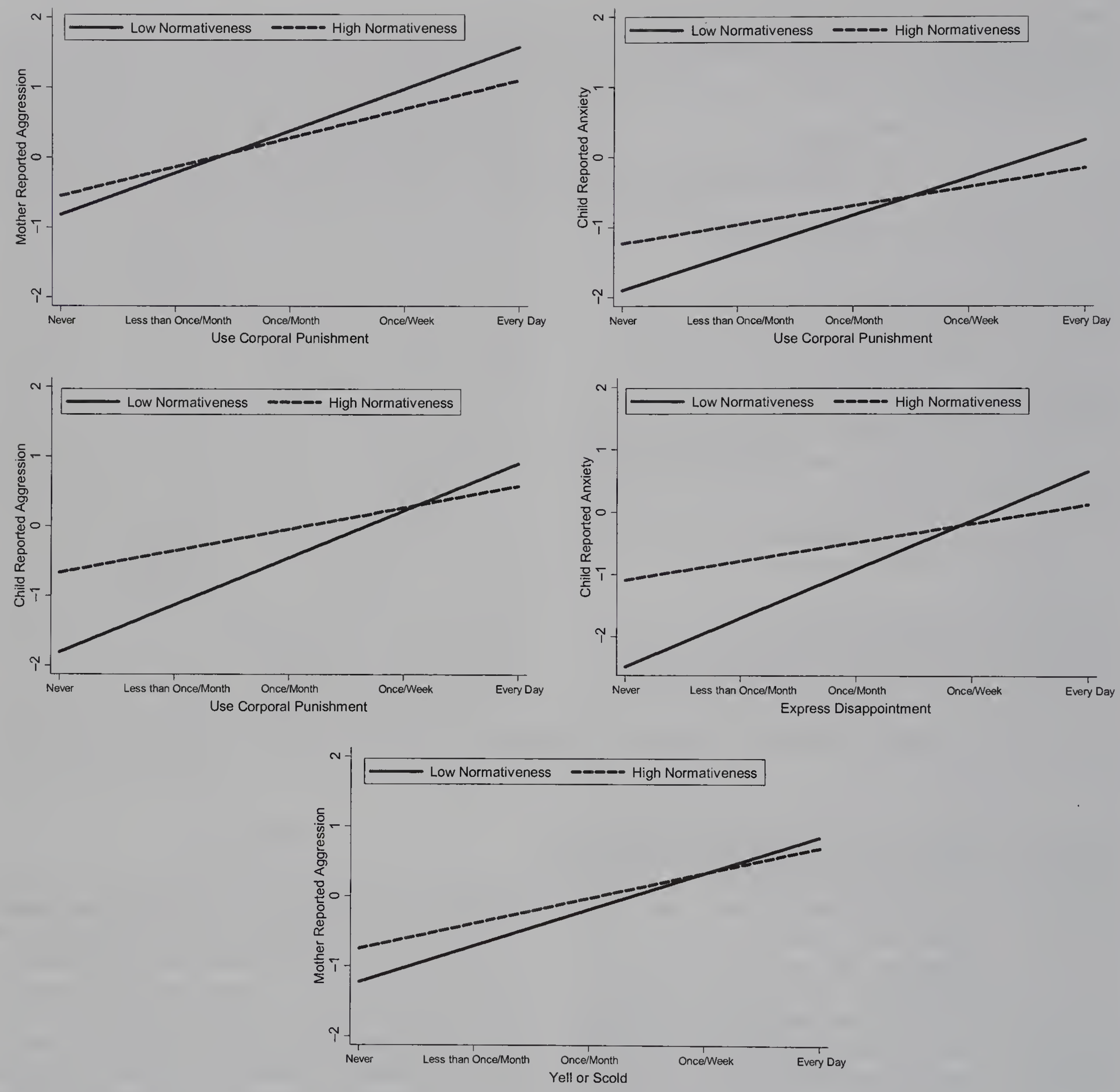


Figure 2. Plots of the links between the frequencies with which mothers use corporal punishment, express disappointment, and yell or scold with children's aggression or anxiety symptoms as moderated by children's perceptions of the normativeness of each discipline technique.

Note. "Low normativeness" and "high normativeness" reflect the frequency with which mothers report using each discipline technique at −1 and +1 *SD*, respectively, from the mean of children's reports of normativeness across the sample.

discipline associated with children's aggressive and anxious behaviors across an international sample made up of families in six countries? The first conclusion to be drawn from our parallel multilevel and fixed effects models that considered all 11 discipline techniques simultaneously is that only three techniques were significantly associated with levels of child aggression and only four techniques were significantly related to levels of child anxiety. In other words, the frequencies with which mothers in this international sample reported using most of the disciplinary techniques were not associated with their children's aggressive or anxious behaviors. The second conclusion is that none of the discipline techniques was associated with less child aggression or less child anxiety. Using corporal punishment, expressing disappointment, and yelling or scolding were each associated with more child aggression, while giving a time-out, using corporal punishment, expressing disappointment, and shaming were associated with more child anxiety.

For the techniques found to be significantly associated with child aggression or anxiety when country was controlled, we added random intercepts to the multilevel models to determine whether there were in fact significant between-country differences in the extent to which that discipline technique was related to child behavior. Of the nine models examined in this way, significant country differences were found only for the association of *express disappointment* with mother-rated aggression; the association was stronger among families in China, Italy, the Philippines, and Thailand than among families in Kenya. However, normativeness does not explain theses differences; normativeness did not moderate the association of *express disappointment* with mother-rated aggression in the final set of analyses. Why this particular discipline technique is more strongly associated with mothers' ratings of child aggression in some countries than in others remains a question for future research.

The cross-sectional nature of the data raises the question of direction of effect. Specifically, does parents' use of these discipline techniques predict more aggression in children, or do aggressive children elicit more discipline? The question of whether such associations are evidence of a parent effect or a child effect has been raised most often with regard to parents' use of corporal punishment (Baumrind, Larzelere, & Cowan, 2002; Larzelere, Kuhn, & Johnson, 2004). It is undoubtedly true in some cases that aggressive children elicit more parental discipline; difficult children have indeed been found to elicit harsh behavior even from nonfamiliar adults (Anderson, Lytton, & Romney, 1986). It is a logical extension of this argument that difficult children should elicit more of *all* forms of discipline from their parents, not just corporal punishment (Larzelere et al., 2004).

Although these data do not permit a rigorous test of this important question (at minimum, longitudinal data would be needed), some patterns of the correlational effects are suggestive of the presence of parent effects on child behavior. First, we did not find that parents use more of all forms of discipline with aggressive children; rather, only 4 of the 11 forms of discipline were significantly associated with child aggression. If aggressive children were eliciting more parent discipline overall, we would have expected them to elicit more of the most frequently reported forms of discipline. However, although mothers reported they *teach about good and bad behavior* and *get child to apologize* most often, they were not using these techniques any more frequently for children they rated as high in aggression. These findings are not consistent with a "child effect" explanation.

Second, although it is not hard to imagine an aggressive child eliciting more discipline from a parent, it is less clear whether a child's symptoms of anxiety would elicit discipline, particularly potentially harsh techniques such as corporal punishment and yelling. Rather, longitudinal research has found that parents typically use less, not more, harsh discipline with anxious and fearful children (Kochanska, Aksan, & Joy, 2007). They do so with good reason: Discipline techniques high in power assertion, including punishment, have been found to substantially undermine moral internalization among children high in fearfulness or anxiety (Kochanska et al., 2007). Children's levels of depression and anxiety have also been found to predict less corporal punishment in particular (Grogan-Kaylor & Otis, 2007). Clearly these analyses will need to be replicated with quasi-experimental or longitudinal data to definitively identify direction of effect, but our findings for both child aggression and child anxiety are more consistent with a parent effect rather than a child effect explanation.

Our decision to use both multilevel and fixed effects analyses exploited the benefits and reduced the costs of using either method alone. The overwhelming consistency of the results across the two sets of models added to our confidence in the robustness of our findings. We encourage future researchers to similarly employ parallel methods when analyzing complex and nested datasets.

Although the associations of discipline with child behavior were almost entirely consistent across cultures when indexed as countries, we did find some evidence that more proximal measures of culture, namely, children's beliefs about the normativeness of discipline within their local communities, did dampen the associations of three forms of discipline with child behavior. In all cases, however, the slopes were positive, such that more frequent use of a particular discipline was associated with greater child behavior problems; even when moderation was present, the association between a discipline technique and a child outcome never reversed direction. This association is also evident in the higher intercepts for the high normativeness group in all of the graphs of the significant moderations. If children's beliefs about normativeness are largely driven by what they see their own parents doing, then it makes sense that the effects of perceived normativeness would be similar in direction and magnitude to the effects of their own parents' use of discipline. It of course may also be that children with higher symptomatology perceive adults in their communities as more punitive. We need to know more about how children form beliefs about normativeness to truly understand these findings.

Mothers' perceptions of normativeness moderated only one of the discipline–child behavior associations. This finding suggests that mothers' perceptions of normativeness and children's perceptions of normativeness may have different functions. Further exploration into how parents develop beliefs about discipline normativeness and into how such beliefs may moderate their own behavior is an exciting avenue for future research.

We had hypothesized that normativeness would moderate the associations of techniques most likely to be perceived as negative events by the child, namely, corporal punishment and yelling, with child behaviors. This was indeed found to be the case. The associations of corporal punishment frequency with child aggression and child anxiety, and of yelling and scolding with child aggression, were less strong when children perceived these techniques to be normative, although again the slope was always positive (more frequent use of the technique associated with a greater number of problem behaviors). We also found the association between expressing disappointment and child-reported anxiety to be moderated by both mothers' and children's perceptions of the normativeness of the technique. In our data, normativeness only played a role for the most negative of the discipline techniques, namely corporal punishment and yelling. Given that the findings for *express disappointment* were similar to these more objectively negative forms of discipline, it appears that when parents express disappointment in their children, children experience that as a distressing event that is linked with more behavior problems the less common it is in the children's community. All other forms of discipline were not affected by normativeness, which suggests to us that the effects of some forms of discipline are robust to community norms. However, given the limited number of significant associations between discipline and child behavior, there is typically no effect of discipline for normativeness to moderate.

Although we worked to ensure the linguistic and cultural equivalence of each discipline technique across the six countries through qualitative interviews and translation and back-translation procedures, it is still possible that there remained some cultural differences in the interpretation of the types of discipline techniques or in the extent to which responses are affected by social desirability bias. An important extension of this work would be to determine whether discipline measures created as indigenous to each country would generate other discipline techniques that are used often in some countries while used rarely or not at all in others (i.e., taking a truly ''emic'' approach; Berry, 1999; Dasen & Mishra, 2000; Harkness & Super, 2002). Additionally, the measure with which we asked parents about their disciplinary techniques focused on individual strategies and not on how parents might combine them; more detailed information, such as diary records (e.g., Chapman & Zahn-Waxler, 1982; Larzelere, Schneider, Larson, & Pike, 1996), could help identify patterns of disciplinary combinations. We also note that the CBCL and YSR subscales used in this article were only modestly internally consistent for the mothers and children in the Indian sample. This was unexpected given that the CBCL has been used successfully with families from the Indian subcontinent (Shenoy et al., 1998).

This study advances our very limited knowledge about parents' use of a wide range of discipline techniques around the world and about links between use of these techniques and children's aggressive and anxious behaviors. We found little evidence that the associations of these discipline techniques with children's behaviors vary by the country in which families live. However, our findings that children's perceptions of the normativeness of discipline were related directly to their behavior and that they moderated the associations between parent discipline and child behavior confirm a potentially important role for children's perceptions of normativeness in understanding how parent discipline impacts child behavior. We encourage more researchers to consider the role of perceived normativeness in their studies of parental discipline and to assess the opinions of those most affected by discipline, the children themselves.

References

Achenbach, T. M. (1991a). *Manual for the Child Behavior Checklist and 1991 profile*. Burlington: University of Vermont, Department of Psychiatry.

Achenbach, T. M. (1991b). *Manual for the Youth Self-Report Form and 1991 profile*. Burlington: University of Vermont, Department of Psychiatry.

Achenbach, T. M. (2004). Cross-cultural perspectives on developmental psychopathology. In U. P. Gielen & J. Roopnarine (Eds.), *Childhood and adolescence: Cross-cultural perspectives and applications* (pp. 411–429). Westport, CT: Praeger.

Allison, P. (2005). *Fixed effects regression methods for longitudinal data using SAS*. Cary, NC: SAS Publishing.

Anderson, K. E., Lytton, H., & Romney, D. M. (1986). Mothers' interactions with normal and conduct-disordered boys: Who affects whom? *Developmental Psychology, 22*, 604–609.

Artigas, J. (1999). Manifestaciones psicológicas de la epilepsia en la infancia. *Revista de Neurologia, 28*, S135–S141.

Baumrind, D., Larzelere, R. E., & Cowan, P. A. (2002). Ordinary physical punishment: Is it harmful? Comment on Gershoff (2002). *Psychological Bulletin, 128*, 580–589.

Berry, J. W. (1999). Emic and etics: A symbiotic conception. *Culture and Psychology, 5*, 165–171.

Bornstein, M. H. (1995). Form and function: Implications for studies of culture and human development. *Culture and Psychology, 1*, 123–137.

Chapman, M., & Zahn-Waxler, C. (1982). Young children's compliance and noncompliance to parental discipline in a natural setting. *International Journal of Behavioral Development, 5*, 81–94.

Collins, W. A., Madsen, S. D., & Susman-Stillman, A. (2002). Parenting during middle childhood. In M. H. Bornstein (Ed.), *Handbook of parenting: Vol. 3. Being and becoming a parent* (2nd ed., pp. 73–101). Mahwah, NJ: Erlbaum.

Crijnen, A. A. M., Achenbach, T. M., & Verhulst, F. C. (1997). Comparisons of problems reported by parents of children in 12 cultures: Total problems, externalizing, and internalizing. *Journal of the American Academy of Child and Adolescent Psychiatry, 36*, 1269–1277.

Dasen, P. R., & Mishra, R. C. (2000). Cross-cultural views on human development in the third millennium. *International Journal of Behavioral Development, 24*, 428–434.

Deater-Deckard, K., & Dodge, K. A. (1997). Externalizing behavior problems and discipline revisited: Nonlinear effects and variation by culture, context, and gender. *Psychological Inquiry, 8*, 161–175.

Deater-Deckard, K., Dodge, K. A., & Sorbring, E. (2005). Cultural differences in the effects of physical punishment. In M. Rutter & M. Tienda (Eds.), *Ethnicity and causal mechanisms* (pp. 204–226). Cambridge, UK: Cambridge University Press.

Dodge, K. A., & Pettit, G. S. (2003). A biopsychosocial model of the development of chronic conduct problems in adolescence. *Developmental Psychology, 39*, 349–371.

Dong, Q., Wang, Y., & Ollendick, T. H. (2002). Consequences of divorce on the adjustment of children in China. *Journal of Clinical Child and Adolescent Psychology, 31*, 101–110.

Ember, C. R., & Ember, M. (2005). Explaining corporal punishment of children: A cross-cultural study. *American Anthropologist, 107*, 609–619.

Florencio, C. (1988). *Nutrition, health and other determinants of academic achievement and school-related behavior of Grades one to six pupils*. Quezon City, Philippines: University of the Philippines.

Gershoff, E. T. (2002). Corporal punishment by parents and associated child behaviors and experiences: A meta-analytic and theoretical review. *Psychological Bulletin, 128*, 539–579.

Giles-Sims, J., Straus, M. A., & Sugarman, D. B. (1995). Child, maternal, and family characteristics associated with spanking. *Family Relations, 44*, 170–176.

Gill, R., & Kang, T. (1995). Relationship of home environment with behavioural problems of pre-school children. *Indian Journal of Psychometry and Education, 26*, 77–82.

Grogan-Kaylor, A., & Otis, M. D. (2007). The predictors of parental use of corporal punishment. *Family Relations, 56*, 80–91.

Grusec, J. E., & Goodnow, J. J. (1994). Impact of parental discipline methods on the child's internalization of values: A reconceptualization of current points of view. *Developmental Psychology, 30*, 4–19.

Guo, L., Leung, P. W. L., Ho, T., Xiong, S., Li, Z., Deng, Z., et al. (2000). A cross-site comparing study on child behavior and emotional problems in Hong Kong and Chengdu. *Chinese Mental Health Journal, 14*, 145–148.

Harkness, S., & Super, C. M. (2002). Culture and parenting. In M. H. Bornstein (Ed.), *Handbook of parenting: Vol. 2. Biology and ecology of parenting* (2nd ed., pp. 253–280). Mahwah, NJ: Erlbaum.

Hoffman, M. L. (1983). Affective and cognitive processes in moral internalization. In E. T. Higgins, D. N. Ruble, & W. W. Hartup (Eds.), *Social cognition and social development* (pp. 236–274). New York: Cambridge University Press.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations* (2nd ed.). Newbury Park, CA: Sage.

Holden, G. W., Miller, P. C., & Harris, S. D. (1999). The instrumental side of corporal punishment: Parents' reported practices and outcome expectancies. *Journal of Marriage and the Family, 61*, 908–919.

Hox, J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Erlbaum.

Kochanska, G., Aksan, N., & Joy, M. E. (2007). Children's fearfulness as a moderator of parenting in early socialization: Two longitudinal studies. *Developmental Psychology, 43*, 222–237.

Korbin, J. E. (2003). Children, childhoods, and violence. *Annual Review of Anthropology, 32*, 431–446.

Lansford, J. E., Chang, L., Dodge, K. A., Malone, P. S., Oburu, P., Palmérus, K., et al. (2005). Physical discipline and children's adjustment: Cultural normativeness as a moderator. *Child Development, 76*, 1234–1246.

Larzelere, R. E., Kuhn, B. R., & Johnson, B. (2004). The intervention selection bias: An unrecognized confound in intervention research. *Psychological Bulletin, 130*, 289–303.

Larzelere, R. E., Schneider, W. N., Larson, D. B., & Pike, P. L. (1996). The effects of discipline responses in delaying toddler misbehavior recurrences. *Child and Family Behavior Therapy, 18*, 35–57.

Lepper, M. R. (1983). Social control processes and the internalization of social values: An attributional perspective. In E. T. Higgins, D. N. Ruble, & W. W. Hartup (Eds.), *Social cognition and social development* (pp. 294–330). New York: Cambridge University Press.

Maas, C. J. M., & Hox, J. J. (2004). Robustness issues in multilevel regression analysis. *Statistica Neerlandica, 58*(2), 127–137.

Mistry, J., Chaudhuri, J. H., & Diez, V. (2003). Ethnotheories of parenting: At the interface between culture and child development. In R. M. Lerner, D. Wertlieb, & F. Jacobs (Eds.), *Handbook of applied developmental science: Vol. 1. Applying developmental science for youth and families: Historical and theoretical foundations* (pp. 233–257). Thousand Oaks, CA: Sage.

Rao, N., McHale, J. P., & Pearson, E. (2003). Links between socialization goals and child-rearing practices in Chinese and Indian mothers. *Infant and Child Development, 12*, 475–492.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

Ripoll-Nuñez, K. J., & Rohner, R. P. (2006). Corporal punishment in cross-cultural perspective: Directions for a research agenda. *Cross-Cultural Research, 40*, 220–249.

Rogoff, B. (2003). *The cultural nature of human development*. New York: Oxford University Press.

Rothbaum, F., Morelli, G., Pott, M., & Liu-Constant, Y. (2000). Immigrant-Chinese and Euro-American parents' physical closeness with young children: Themes of family relatedness. *Journal of Family Psychology, 14*, 34–348.

Save the Children. (2008). *The Child Development Index: Holding governments to account for children's well-being*. Retrieved December 11, 2008, from http://www.savethechildren.org.uk/en/docs/child-development-index.pd

Sears, R. R., Maccoby, E. E., & Levin, H. (1957). *Patterns of child rearing*. New York: Harper & Row.

Shenoy, J., Kapur, M., & Kaliaperumal, V. G. (1998). Psychological disturbance among 5- to 8-year-old school children: A study from India. *Social Psychiatry and Psychiatric Epidemiology, 33*, 66–73.

Socolar, R. R. S., Savage, E., & Evans, H. (2007). A longitudinal study of parental discipline of young children. *Southern Medical Journal, 100*, 472–477.

Stata Corporation. (2005). *Longitudinal/panel data reference manual*. College Station, TX: Stata Press.

Straus, M. A. (1979). Measuring intrafamily conflict and violence: The Conflict Tactics Scale. *Journal of Marriage and the Family, 41*, 75–88.

Straus, M. A., & Mouradian, V. E. (1998). Impulsive corporal punishment by mothers and antisocial behavior and impulsiveness of children. *Behavioral Sciences and the Law, 16*, 353–374.

Straus, M. A., & Stewart, J. H. (1999). Corporal punishment by American parents: National data on prevalence, chronicity, severity, and duration, in relation to child and family characteristics. *Clinical Child and Family Psychology Review, 2*, 55–70.

Super, C. M., & Harkness, S. (1986). The developmental niche: A conceptualization at the interface of child and culture. *International Journal of Behavioural Development, 9*, 545–569.

Weisz, J. R., Sigman, M., Weiss, B., & Mosk, J. (1993). Parent reports of behavioral and emotional problems among children in Kenya, Thailand, and the United States. *Child Development, 64*, 98–109.

Weisz, J. R., Suwanlert, S., Chaiyasit, W., Weiss, B., Achenbach, T. M., & Eastman, K. L. (1993). Behavioral and emotional problems among Thai and American adolescents: Parent reports for ages 12–16. *Journal of Abnormal Psychology, 102*, 395–403.

World Bank. (2008, October). *Revised estimates for PPP GNI per capita income and PPP GDP*. Retrieved December 11, 2008, from http://go.worldbank.org/B5PYF93QF0

Child Development, March/April 2010, Volume 81, Number 2, Pages 503–516

Coercive and Prosocial Fathering, Antisocial Personality, and Growth in Children's Postdivorce Noncompliance

David Scott DeGarmo
Oregon Social Learning Center

To better understand quantity and quality of divorced father contact, a weighted county sample of 230 divorced fathers with a child aged 4–11 years was employed to test whether fathers' antisocial personality (ASP) moderated effects of monthly contact with children in predicting children's observed noncompliance. Eighteen-month latent growth models obtained significant individual differences in levels of noncompliance and growth rates. ASP significantly moderated beneficial impact of fathers' monthly contact. Fathers' observed parenting practices significantly predicted noncompliance levels but not growth. Parenting did not account for the effect of Contact × ASP, suggesting both environmental and potentially genetic influences on child adjustment. Findings were robust across boys and girls and age levels. Implications for preventive intervention are discussed.

Recent representative findings are showing that fathers' involvement is beneficial for children following divorce; however, this influence is not beneficial if father contact is influenced or moderated by fathers' antisocial behaviors or antisocial personality (ASP; Coley & Medeiros, 2007; Jaffee, Moffitt, Caspi, & Taylor, 2003; King & Sobolewski, 2006). The existing research testing these effects has focused mainly on adolescents and little research has been carried out focusing on father's actual parenting behaviors in addition to fathers' amount of contact. The present study tests whether fathers' monthly contact following divorce is moderated by antisociality for younger children and whether parenting practices account for this relationship. Multiple methods were used including direct observation of fathering behaviors and child noncompliance to better address our understanding of the quantity and quality of divorced father contact.

Father Involvement and Child Adjustment Following Divorce

A generative fathering perspective focuses on a positive growth-promoting model seeking to transcend the deficit model of men in families (Brotherson, Dollahite, & Hawkins, 2005; Hawkins & Dollahite, 1997). This perspective is particularly relevant for divorced fathers (Pasley & Minton, 1997) given that for some, the role of father may increase in importance and fathers may become more involved with their children, while for others, factors such as increased role strains, ambiguities over parenting, conflict with the former spouse, and lack of institutional and personal supports can contribute to decreases in father contact and postdivorce parenting (Braver et al., 1993; Stephens, 1996).

Contributing to the pervasive dichotomy of the good father or bad father is research on fathers' influence on children following divorce or separation. Studies on resident and nonresident fathers (divorced and never married) have produced mixed findings showing that fathers' involvement can contribute both positively and negatively to child development. A generative perspective is supported from studies showing that quality father involvement of nonresident fathers has beneficial impacts children's adjustment (Simons, Whitbeck, Beaman, & Conger, 1994). For example, in a meta-analysis of younger and older children, Amato and Gilbreth (1999) found that nonresidential father involvement and authoritative parenting (e.g., instructive, noncoercive parenting) were associated with higher levels of children's academic success and lower levels of internalizing and externalizing. More recently, King and Sobolewski (2006) found

---

Correspondence concerning this article should be addressed to David Scott DeGarmo, Oregon Social Learning Center, 10 Shelton McMurphey Boulevard, Eugene, OR 97401. Electronic mail may be sent to davidd@oslc.org.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8102-0008

that father involvement and father–child relationship quality in the National Survey of Households and Families predicted lower levels of internalizing problems for adolescents with high father involvement compared to adolescents with low-involved fathers. Similarly, data from the Three-City Representative Study of Low-Income Fathers also found that greater nonresident father involvement predicted subsequent decreases in adolescent delinquency, particularly for high-delinquent adolescents (Coley & Medeiros, 2007). Similar results were found for father involvement in the United Kingdom for families at risk for children's conduct problems (Dunn, Cheng, O'Connor, & Bridges, 2004).

In contrast, father involvement can have a detrimental impact on children's well-being. Key findings from an epidemiological twin study found that the negative impact of fathers' contact and parenting on children's problem behaviors was conditioned by fathers' own antisocial behaviors. Jaffee et al. (2003) found that the more time children lived with fathers, the fewer conduct problems children had; however, this finding was moderated by antisocial behaviors of fathers measured from interview data of lifetime prevalence of antisocial behaviors (e.g., stealing, fighting, antisocial beliefs). Jaffee et al. reported a "double whammy" of genes and environments for antisocial fathers influencing child behaviors. They found that antisocial behaviors are highly heritable; at the same time, however, effects of father ASP on children's behavior problems remained significant controlling for (a) any genetic influence of the father, (b) influence of resident nonbiological father figures, and (c) influence of maternal antisocial characteristics, thus suggesting the importance of parenting. Also supporting an environmental effect, Simons et al. (1994) found main effects showing that nonresident divorced fathers' positive reinforcement and noncoercive discipline were associated with lower levels of adolescent conduct problems. Fathering effects were also independent of the mothers' parenting and cocustody conflict in the sample of coparenting couples. These findings showing the impact of antisocial fathers and coercive discipline are consistent with the coercion model of family process (Patterson, 1982).

Together these studies suggest that the impact of nonresident father involvement is beneficial on children's problem behaviors in the absence of fathers' antisocial behaviors or in the absence of coercive parenting practices. With the exception of the Simons et al. (1994) study, very little research has been carried out on divorced fathers' actual fathering behaviors or parenting practices. Until recent federal fatherhood initiatives, the majority of prior studies have relied on single sources of data or mother reports of father involvement (Parke et al., 2004), chiefly, the frequency or amount of contact and child support (Dunn et al., 2004; Pasley & Braver, 2004). Furthermore, studies on divorced fathers' parenting are still relatively small in number compared to studies of fathers in general. Research has also focused mainly on the period of adolescence. The present study attempts to add to our understanding of fathering impacts on child development by operationalizing father contact as a continuum and father's effective parenting as a continuum to test generative and coercion model perspectives in a sample of recently divorced fathers and their young pre- and elementary-school-aged children, a group shown to be more vulnerable to the effects of divorce (Hetherington & Clingempeel, 1992).

Theoretical Model of Fathers' Parenting and Growth in Children's Problem Behaviors

Several studies have shown how the marked and significant changes in parenting routines, residence, and social roles during marital separation, divorce, and repartnering directly interfere with effective parenting and subsequent child adjustment (Capaldi & Patterson, 1991; Hetherington & Clingempeel, 1992; Ronald L. Simons and Associates, 1996; Simons, Beaman, Conger, & Chao, 1993). The present study employs Patterson's (1982) coercion model specifying how parents' antisocial characteristics are linked to coercive parenting behaviors that are the mechanism reinforcing the development of and growth in children' antisocial behaviors. The theoretical model is also more broadly known as the social interactional–social learning model of child development (Patterson, 1982, 2005; Reid, Patterson, & Snyder, 2002) because it focuses on the influence of the family social environment in shaping overlearned patterns of behavior that can generalize across social settings for the child. Parenting practices are seen as the key agent of child socialization. Parent–child interaction patterns (i.e., *a social interactional perspective* of parenting) are the behavioral mechanism reinforcing the child's likelihood of engaging in future prosocial or antisocial behaviors (i.e., *the social learning perspective* of child development). A child's interpersonal style and behavioral repertoire learned within the family is presumed to carry over to his or her interactions with peers and adults in other settings.

More specifically, a coercion model focuses on how harsh punitive and ineffective discipline is associated with growth in children's problem behaviors. There are several ways in which a child's problem behaviors can be reinforced. A classic example is a child requesting candy in the grocery store. When a parent says ''no'' to the child's request, a child may throw a tantrum that is aversive to the parent. An unskilled parent may reward the child with candy to terminate the aversive tantrum, thus reinforcing the likelihood of future tantrums and a child learning to be rewarded. Similarly, an unskilled parent may use hitting, yelling, and harsh punishment to discipline a child to extinguish an aversive child behavior when a child misbehaves. Coercive discipline works in the short run; however, a child can learn that coercion among family members can be used to terminate aversive behaviors of other family members. A child may then learn to apply coercion in relationships with peers and other adults. In the absence of skilled parenting, a child may progress from trivial displays of aversive behaviors to growth in aggression and future delinquency (Patterson & Yoerger, 1999).

The coercion model is relevant for understanding fathering. Studies have shown that antisocial behaviors of fathers and coercive parenting are associated with increased child and adolescent behavior problems over time (Dishion, Owen, & Bullock, 2004; Jaffee et al., 2003). Relative to mothers, one study reported that fathers' inept discipline can explain twice the variance in children's problem behaviors (Patterson & Dishion, 1988). Dishion et al. (2004) employed the Oregon Youth Study (OYS) to test competing models of a direct effect of parenting (a family management model) or a direct effect of father's deviance (a cultural deviance model) on the early adolescent deviant peer association. They found a direct link between father antisocial behaviors and the selection of deviant peers, suggesting that fathers have an important leadership and modeling role in the socialization of sons with respect to norms, values, and deviance, with antisocial fathers potentially encouraging it. Dishion et al. also found significant associations from father antisocial characteristics to less effective parenting and boys' antisocial behaviors in early childhood as shown previously in the OYS.

The broader social interactional–social learning model also specifies a methodological paradigm using multiple-method assessments including observation to more reliably measure parent–child interaction patterns. Most recently, this model has also focused on evaluation of parenting training outcomes in randomized intervention trials (Patterson, 2005). For fathering, an experimental evaluation of parent training for a sample of 110 recently married stepfather families produced medium effect sizes, demonstrating improvements in observational-based measures of stepfathers' prosocial parenting and coercive parenting at 6 and 12 months postintervention follow-ups (DeGarmo & Forgatch, 2007). These improvements in stepfathering predicted reductions in children's reported internalizing and reductions in observed noncompliance, independent of mothers' parenting. In another report of the same study, the intervention produced significant improvements in mother–father coparenting behaviors, which were associated with reductions in mother-, stepfather-, and teacher-reported problem behaviors at the 2-year follow-up (Forgatch, DeGarmo, & Beldavs, 2005).

The present study based on the social interactional–social learning model of child development is not an experimental trial but is the first effort to test the hypothesized coercion model with a sample of recently divorced fathers. Like the stepfather intervention study discussed earlier, the focus here is on children's noncompliance, a primary marker in young children for the development of later conduct disorder (Martinez & Forgatch, 2001; Vuchinich, Hetherington, Vuchinich, & Clingempeel, 1991). The present study was designed to examine fathering effects across a range of custodial responsibility and father–child contact. In the first report of this data set, the amount of contact and time spent with children varied according to custody as expected: Full-custody fathers reported more time with children than shared-custody fathers, and each respectively more time than no-custody fathers. However, no differences were obtained by custody on coercive parenting practices (DeGarmo, Patras, & Eap, 2008). That is, the amount of time varied by custody but not the level of coercive fathering. Because no-custody fathers can have regular contact with children, and therefore potential impact, the social interactional parenting model is tested using father contact as a continuum rather than court-defined custody status.

Present Study Hypotheses

The following hypotheses were developed based on the literature above:

1. Divorced fathers' monthly contact is expected to have a beneficial impact preventing

growth in children's noncompliance over time but is hypothesized to be moderated by level of fathers' ASP, such that contact by fathers with high ASP is expected to be detrimental and contact by fathers with low ASP is expected to be beneficial.

2. Fathers' coercive parenting practices are expected to predict children's noncompliance over time and will account for effects of fathers' ASP and levels of contact.

Method

Participants

A total of 230 fathers were recruited to participate in a study focusing on father and child adjustment to divorce, the Oregon Divorced Father Study (ODFS). A unique aspect of the study was to evaluate adjustment for divorced fathers, including a range of father–child contact including custodial and noncustodial fathers. Therefore, all divorced fathers with children in the focal age range were eligible for selection within the county sampling frame. Fathers were recruited through public court records. Eligible fathers had a focal child between the ages of 4 and 11 (boy or girl) and had obtained a divorce decree within the prior 24 months. Court records were screened for children within the targeted age range. We randomly selected the focal child if there was more than one eligible child. If the father could not enroll that child, we randomly selected the next eligible child for potential participation until we either enrolled a child or exhausted all eligible children.

After selecting eligible records, fathers were invited to participate with their child through a recruitment letter explaining the study and how they were selected. Among the 230 enrolled fathers, 180 (78%) chose and were able to enroll the focal child. For comparison of nonparticipants, custody status was defined as legal custody reported in the court records. We found that 92% of full-custody and 95% of shared-custody fathers enrolled the targeted focal child, whereas 41% of no-custody fathers enrolled their focal child. Therefore, not all full- and shared-custody fathers had their children participate in the center assessments. However, all fathers, including no custody, filled out questionnaires and interview data on their children's behavior and their own parenting practices. The large majority of fathers (96%) reported contact with the focal child with in-person visitation, e-mail, telephone, or written correspondence.

Fathers' age ranged from 22.9 to 63.4 ($M = 37.8$, $SD = 7.7$), education from 1 (< *8th grade*) to 13 (*advanced doctorate*; $M = 7.2$, $SD = 2.9$), and income from 1 (< *$5K*) to 10 (> *$100K*; $M = 5.4$, $SD = 2.2$). Twenty-three percent of the fathers were below federal guidelines for poverty thresholds. Children's age ranged from 4.0 to 12.0 ($M = 7.6$, $SD = 2.0$). Forty-seven percent of the enrolled children were girls. Thirteen percent of the fathers identified themselves as a racial minority (5% Hispanic, 1% African American, 1% Asian American, 1% Native American, and 5% multiracial) and 17% of their children (1% Hispanic, > 1% African American, > 1% Asian American, 1% Native American, and 14% multiracial), reflective of county demographics. The county sampling frame consisted of 37% no-, 54% shared-, and 10% full-custody fathers. The ODFS comprised 32% no-, 53% shared-, and 14% full-custody fathers with an overall participation rate of 40% (35% no-, 41% shared-, and 55% full-custody fathers).

Neighborhood demographics were collected for eligible participants and nonparticipants. There were no differences between participants and nonparticipants when comparing neighborhood characteristics from census tract and geo-coded police call response data near the father's address (e.g., unemployment rates, proportion homeownership, poverty rates, racial makeup, and police call frequency and severity ratings). To generalize beyond the community level to a county level, sample weights were obtained as correction factors for selection bias based on county proportions of custody. In short, the sampling weights adjust for differential participation, eligibility, and location rates by custody and potential selection bias. Details are provided in DeGarmo et al. (2008).

Data Collection and Assessment Procedures

Data were collected during a father–child visit at each wave. Each center visit took approximately 2.5 hr. All participants were provided child care, transportation, and a meal if requested (delivered local take out). Father and child were each paid approximately $25 an hour for their time. Data were from paper-and-pencil questionnaires, face-to-face interviews, self-administered and computer questionnaires, and observational coding of father–child interaction. Observational data were obtained from a total of 24 min of videotaped interaction scored across four structured interaction tasks during the father–child visit. The tasks were a refreshment task lasting 5 min, a problem-solving discussion focusing on a parenting issue lasting 7 min, a play task with the father lasting 7 min, and an academically challenging teaching task lasting 5 min. The refreshment task was a "warm up" in which father and child were asked to take a break from interviews and filling out questionnaires. Each of the problem-solving discussions focused on current father-selected topics from a checklist of common parent–child issues. They were asked to resolve issues rated as "hottest." In the teaching task, fathers were asked to work with the child on a typical homework problem. Children were provided a homework problem that was one grade level beyond the child's current grade.

Trained observers scored father–child interaction with the Family and Peer Process Code (FPP; Stubbs, Crosby, Forgatch, & Capaldi, 1998) obtaining data on discrete positive, negative, and neutral behaviors in real-time coding along with data on the initiator, recipient, sequence, and affect. Fifteen percent of videotapes were randomly selected for blind reliability checks. For all content and affect codes, two coefficients of intercoder reliability were assessed: percent agreement and Cohen's Kappa, an indicator of agreement above chance. Reliabilities are reported next (FPP manual available at http://www.oslc.org/resources/codingsystems.html).

Father ASP

Fathers' antisocial characteristics were measured using two self-report instruments. The first indicator was the 20-item Acting Out subscale of the Minnesota Multiphasic Personality Inventory–Test–Retest Index (MMPI–TRI; Swanson, Templer, Streiner, Reynolds, & Miller, 1995). The MMPI–TRI is a shortened version of the MMPI and MMPI–2 self-report personality inventories. The MMPI–TRI has demonstrated construct validity with the full MMPI and has been normed with inmate and psychiatric populations (Swanson et al., 1995). Twenty acting-out items evidencing content validity and the highest point-biserial item-total correlations were retained for the shortened MMPI. Statements were rated true or false and then summed. Items included: was suspended from school one or more times, when people do me wrong I feel I should pay them back, when I was young I stole things, at times I feel like picking a fist fight with someone, I can easily make people afraid of me and sometimes do it for fun, and so on. Cronbach's alpha for the ODFS was .82.

The second indicator was the Agreeableness subscale (NEO–A) from the NEO–Five Factor Inventory (NEO–FFI; Costa & McCrae, 1995). The NEO is a standard personality inventory with alpha coefficients of internal consistency ranging from .89 to .92 for extroversion, openness, and agreeableness in validation studies of men and women (Costa & McCrae, 1995). The NEO–A was chosen because it has been validated with the DSM–IV manual as a marker of ASP for men scoring low on the NEO–A (Costa & McCrae, 1992). Agreeableness was a 12-item subscale rated from 5 (*disagree*) to 1 (*strongly agree*). Items included: I often get into arguments with my family and coworkers, if I don't like people I let them know it, if necessary I am willing to manipulate people to get what I want, I try to be courteous of everyone I meet, most people I know like me, and so on. Items were rescaled 0–4 and summed to obtain norming scores. Data from the norming sample of 500 adult men produced a mean of 31.93 ($SD = 5.03$) for the NEO–A (Costa & McCrae, 1992). ODFS fathers revealed no significant difference in mean or variance ($M = 31.74$, $SD = 5.31$, $t = 0.43$). Cronbach's alpha for the ODFS was .82 for the NEO–A.

Father–Child Monthly Contact

The central focus of the present analyses is fathers' monthly contact with child indexed by day and night visits. Contact was measured from a computerized interview with the father adapted from Braver et al. (1993) and was computed as the mean total number of weekday and weekend days per month of contacts with the child and the number of overnight stays and visits during the typical school year. Time-use measures were not collected; therefore, each visit may have varied from father to father, and monthly contact was therefore a discrete time index of involvement. Monthly contact is confounded with custody status, to illustrate variation within custody, Figure 1 plots the histogram distributions of monthly contact by custody and provides the boxplot distribution of variation for the total sample.

Children's Observed Noncompliance

Noncompliance was the rate per minute (rpm) of a noncomply scored across all father–child interaction tasks. FPP coders first coded any directive or command provided by the father. A toggle code then records whether the child was compliant or cooperative or not. Compliance is defined as the act of clearly obeying a father's request or command and noncompliance the act of clearly disobeying a father's request or command. Only one compliance behavior (either comply or noncomply) may be entered in response to each directive. Hypothetical,

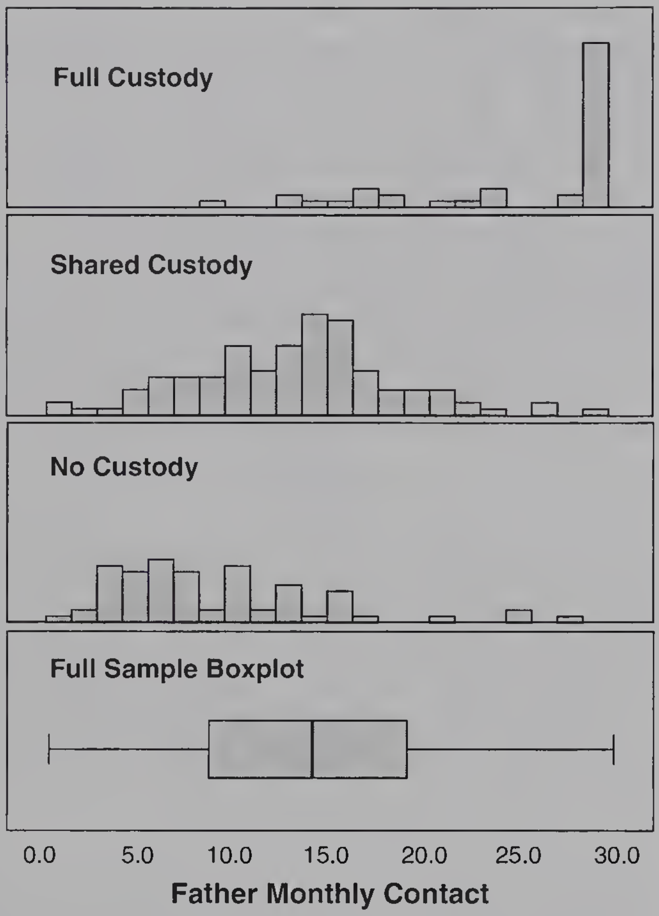


Figure 1. Histogram distribution of index for fathers' monthly contact with child by custody status and boxplot of monthly contact for full sample.

ambiguous, or unclear responses to a directive are not coded. For father command or directives, the percent agreement ranged from .84 to .86 from baseline to the 18-month follow-up. Kappas ranged from .85 to .86. For noncompliance, percent agreements ranged from .66 to .77 over time with Kappas of .75–.77.

Fathers' Parenting Practices

Fathers' parenting was assessed using previously validated indicators assessing "coercive" and "prosocial" fathering using multiple sources of data (DeGarmo & Forgatch, 2007; DeGarmo et al., 2008) including self-report, discrete or microsocial behavior codes from the FPP coding system, and coder Likert-type ratings of parenting. These measures of the latent variable were operationalized with four indicators: ratings of coercive harsh discipline during the structured interaction tasks, coded rpm of verbal and physical aggression of the father directed to the child, self-reported harsh discipline, and ratings of prosocial parenting during the interaction tasks.

Harsh discipline ratings were a five-item scale of the coder's impression of discipline after scoring FPP discrete behaviors. Rated from 1 (*very untrue*) to 5 (*very true*), items included overly strict, authoritarian, expressed hostility during discipline, used nagging, hovered too closely, and used inappropriate discipline. Cronbach's α was .86.

Father total aversives from the FPP coding system included cluster scores of father-initiated behaviors directed to the child that included physical and verbal aggression codes (e.g., physical aggression, verbal aggression, negative tease, negative interpersonal, physical attack, verbal attack; coding system history and details are available for download on the Internet). The FPP percent agreement for reliability checks for content codes were .87, .87, and .89 for content behavior codes, and .95, .96, and .96 for affect codes. Cohen's Kappas were .78, .79, and .81 for content codes over time, and .80, .78, and .76, respectively, for affect codes.

Self-report harsh discipline was a self-administered questionnaire including a five-item summative index rated from 1 (*never or almost never*) to 5 (*always or almost always*) in which fathers responded to the question "When [focal child] misbehaves, how often do you . . . ?" Rated items were: raise your voice/scold, yell, spank on bottom, slap, hit. Cronbach's alpha was .85.

Prosocial parenting was the mean of two scales from coder ratings of positive involvement and skill encouragement. *Positive involvement* was obtained from 14 items rated after each father–child interaction task. Items included ratings on how much the parent treated the child with warmth, empathy, affection, and respect, maintained good eye contact and interactive posture, and so on. Cronbach's alpha was .94.

Skill encouragement was based on ratings of fathers' ability to promote children's skill development through contingent encouragement and scaffolding strategies observed during the teaching and construction play tasks. In both tasks, the child is given challenging problems and the father is asked to assist. The scale includes 11 items such as: breaks task into manageable steps, reinforces success, prompts, corrects appropriately. Cronbach's alpha was .92. Both prosocial scale scores were rescaled from 0 to 1 and averaged to form a composite prosocial parenting indicator.

Control Variables

Age of the child in years computed from date of assessment and date of birth. Sex of child coded as 0 for girls and 1 for boys.

Analytic Strategy

The main study hypotheses were tested with structural equation modeling (SEM) specified as latent variable growth models in the MPlus5 program (Muthén & Muthén, 2007). MPlus provides the ability to apply sample weights using the MLR option, or maximum likelihood estimation with robust standard errors. MLR standard errors and chi-square model test are robust to non-normality. Latent growth models (LGM) provide advantages for modeling developmental changes over time and are a special case of multilevel modeling in the SEM framework in which repeated measure outcomes at Level 1 are nested within individuals at Level 2.

Also known as random intercepts models, LGM estimates the individual differences or variation in the levels of noncompliance (random intercepts) as well as the increases or decreases in child outcomes (random slopes) for each individual at Level 1. The individual intercepts and slopes are then summarized as latent variable factor components at Level 2 representing the sample means and slopes. The present LGM analyses were modeled as two factors, the average level intercepts over time and as linear growth slopes (Biesanz, Deeb-Sossa, Papadakis, Bollen, & Curran, 2004). This is obtained by fixing the three random intercept factor loadings at 1, respectively, for baseline, 9, and 18 months, and by specifying chronometric time weights for the slope factor at −1, 0, and 1. Data were modeled using full-information maximum likelihood (FIML), which uses all available information from the observed data in handling missing data. FIML estimates are computed by maximizing the likelihood of a missing value based on observed values in the data. For any individual who has baseline data only and no follow-up data, such individuals contribute nothing to the likelihood of estimates and are effectively excluded from change analyses (Brown et al., 2008). Compared to mean-imputation, listwise, or pairwise models, FIML provides more statistically reliable standard errors (Wothke, 2000).

To test whether father ASP moderated father contact, variables were entered as centered first-order terms and centered cross products of ASP and monthly contact using regression approaches for testing interactions (Cohen, Cohen, West, & Aiken, 2003). In addition, to test latent variable interactions of ASP, two specific approaches were used. First, ASP factors scores were saved and used to compute a centered cross product of Monthly Contact × ASP. The second method used latent variable interaction modeling in MPlus using maximum likelihood and a numerical integration algorithm (Muthén & Muthén, 2007). Standardized solutions and explained variance are not available for the latent variable interaction approach; therefore, model fit and standardized solutions were reported next using centered cross products, but findings were tested using both methods of assessing interaction effects.

Results

Eighty-four percent of the families were retained at 9 months and 82% at 18 months. Attrition analyses revealed no significant differences between children in families remaining in the study compared with those lost to follow-up by 18 months when comparing the baseline rpm of child noncompliance. Child noncompliance scores were log transformed for the multivariate analyses, although the MLR estimates are robust to non-normality. Descriptives of study variables are provided in Table 1.

The first step in the LGM analysis was to estimate the linear growth in child noncompliance. Known as the unconditional model, the LGM analyses describe the pattern of change for the developmental outcomes. The average level intercept factor represents the continuum of low noncompliance levels to chronically high levels averaged over time. Average rpm levels of noncompliance were different from zero ($M = 0.08$, $p < .001$) for the sample and the variance component was different from zero ($\sigma^2 = 0.005$, $p < .001$), indicating significant individual differences among children's levels of noncompliance.

The linear growth slope or change factor obtained significant variance, meaning there were individual differences in the increases and decrease trajectories of noncompliance ($\sigma^2 = 0.004$, $p < .001$); however, the sample mean was not different from zero. This meant that the sample was characterized by a range of children increasing in noncompliance and children decreasing in noncompliance. The chi-square minimization fit indicated a linear growth model provided good fit to the repeated measures noncompliance data, weighted MLR growth model fit, $\chi^2(1) = 2.26$, $p = .14$; comparative fit index = .95, standardized root mean square residual = .02. Results of the unconditional noncompliance LGM model are provided in Table 2.

The next step of analyses tested Hypotheses 1 and 2. The first model specified father ASP, father contact, and the centered cross-product term as

Table 1

Sample Descriptives for Observed Variables in Latent Growth Analyses

	Min.	Max.	*M*	*SD*
Child noncompliance				
Observed rpm T1	0.00	0.89	0.09	0.13
Observed rpm T2	0.00	0.97	0.10	0.16
Observed rpm T3	0.00	0.58	0.08	0.12
Natural log rpm T1	0.00	0.64	0.08	0.11
Natural log rpm T2	0.00	0.68	0.09	0.12
Natural log rpm T3	0.00	0.46	0.07	0.10
Fathers' parenting				
Observed harsh discipline ratings	0.00	0.95	0.11	0.16
Observed total aversive to child	0.00	0.48	0.06	0.07
Self-reported harsh discipline	5.00	22.00	10.61	2.64
Prosocial parenting ratings	0.59	0.99	0.84	0.08
Predictors				
Father acting out—MMPI–TRI	0.00	19.00	4.94	3.77
Father agreeableness—NEO–FFI	14.00	44.00	31.75	5.29
Father–child day and overnights per month	0.00	30.00	14.16	7.92
Sex of child (boy)	0.00	1.00	0.53	0.50
Child age	3.99	12.00	7.66	1.98

Note. T1 = baseline Time 1; T2 = 9-month follow-up; T3 = 18-month follow-up; rpm = rate per minute; MMPI–TRI = Minnesota Multiphasic Personality Inventory–Test–Retest Index; NEO–FFI = NEO–Five Factor Inventory.

Table 2

Latent Variable Means, Variances, and Standard Errors for 18-Month Unconditional Growth Model of Child Noncompliance

Estimated variable factor	*M*	*SE*	Variance	*SE*
Average level intercept	0.075***	0.006	0.005***	0.001
Linear growth slope	−0.004	0.005	0.004***	0.001

Note. Weighted MLR growth model fit, $\chi^2(1) = 2.26$, $p = .14$; comparative fit index = .95, standardized root mean square residual = .02. MLR = maximum likelihood with robust standard errors.
***$p < .001$.

predictors of levels and growth in child noncompliance. Findings are displayed in Figure 2 in the form of standardized paths. For clarity, only significant paths from the control variables are displayed, and significant paths from the hypothesized constructs are shown in bold. Results supported Hypothesis 1, indicating that father ASP moderated the effect of monthly contact in predicting both the average levels over time ($\beta = .31$, $p < .001$) and linear growth rates ($\beta = .30$, $p < .05$). The main effect of fathers' monthly contact was marginally associated with decreases in noncompliance over time at the $p = .06$ level ($\beta = -.19$, $p < .062$), which is interpreted as the effect of contact at average levels of ASP in the presence of a significant interaction (Cohen et al., 2003). Therefore, ASP conditioned the impact of fathers' contact on children's noncompliance. Older children displayed lower levels of noncompliance relative to younger children ($\beta = -.28$, $p < .001$), and sex of the child was not associated with noncompliance factors. The overall model predicted 39% of the variance in average levels and 11% of the variance in growth rates.

The next model specified linkages from father ASP to noncompliance through fathers' coercive parenting practices. It was expected that higher levels of coercive parenting would be associated with higher levels of child noncompliance and growth in noncompliance. Furthermore, it was predicted that coercive parenting practices would account for the interaction effect of Contact × ASP. Results are displayed in Figure 3. Hypothesis 2 was partially supported. Father ASP predicted higher levels of coercive parenting practices ($\beta = .42$, $p < .01$), and fathers' coercive parenting in turn predicted higher levels of noncompliance ($\beta = .43$, $p < .01$). Parenting practices explained an additional 14% of the variance noncompliance levels. Contrary to expectations, coercive parenting did not predict growth rates. In addition, parenting did not account for the Contact × ASP interaction effect on noncompliance. Therefore, ASP was a significant moderator of fathers' monthly contact in predicting noncompliance following divorce and ASP was a contributor indirectly through coercive parenting practices.

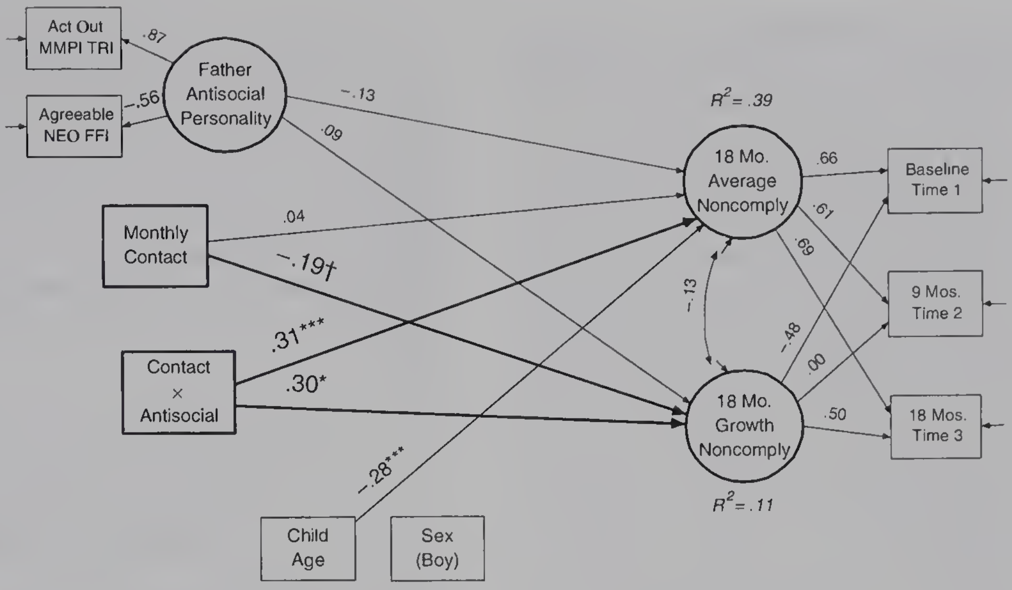

Act Out MMPI TRI
.87
Agreeable NEO FFI
–.56
Father Antisocial Personality
–.13
.09
Monthly Contact
.04
–.19†
Contact × Antisocial
.31***
.30*
–.28***
Child Age
Sex (Boy)
R² = .39
18 Mo. Average Noncomply
–.13
18 Mo. Growth Noncomply
R² = .11
.66
.61
.69
–.48
.00
.50
Baseline Time 1
9 Mos. Time 2
18 Mos. Time 3

Figure 2. Weighted MLR growth model specifying father antisocial personality and monthly contact as predictors of rate per minute of observed child noncompliance over 18 months.
Note. Paths are standardized coefficients. $\chi^2(12) = 21.02$, $p = .44$, $\chi^2/df = 1.75$, standardized root mean square residual = .03, comparative fit index = 1.00. MLR = maximum likelihood with robust standard errors.
$^{\dagger}p < .06$. $^{*}p < .05$. $^{***}p < .001$.

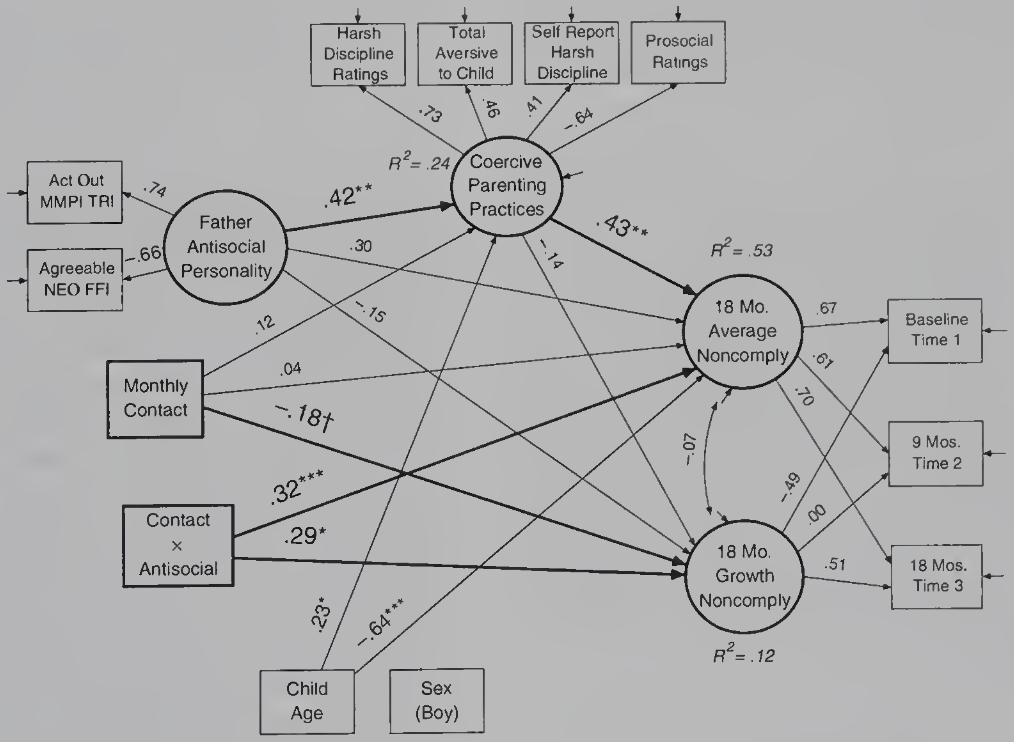

Harsh Discipline Ratings
Total Aversive to Child
Self Report Harsh Discipline
Prosocial Ratings
.73
.46
.41
–.64
R² = .24
Coercive Parenting Practices
Act Out MMPI TRI
.74
Agreeable NEO FFI
–.66
Father Antisocial Personality
.42**
.30
–.15
.12
.43**
–.14
Monthly Contact
.04
–.18†
Contact × Antisocial
.32***
.29*
.23*
–.64***
Child Age
Sex (Boy)
R² = .53
18 Mo. Average Noncomply
–.07
18 Mo. Growth Noncomply
R² = .12
.67
.61
.70
–.49
.00
.51
Baseline Time 1
9 Mos. Time 2
18 Mos. Time 3

Figure 3. Weighted MLR growth model entering fathers' parenting as a predictor of rate per minute of observed child noncompliance over 18 months.
Note. Paths are standardized coefficients. $\chi^2(44) = 99.60$, $p = .000$, $\chi^2/df = 2.26$, standardized root mean square residual = .07, comparative fit index = .82. MLR = maximum likelihood with robust standard errors.
$^{\dagger}p < .06$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Exploratory Models of Gender and Age

Three related approaches were used to explore potential gender and age variation in the main findings. These included (a) multiple group SEMs for boys versus girls and youngers (younger than 8) and olders (older than 8), (b) computed interaction terms, and (c) latent variable interaction models. No moderating effects were obtained comparing boys versus girls as multiple group SEMS. For age, a multiple group split of youngers ($n = 133$) and olders ($n = 97$) would not converge for the older group due to reduced n and variance in the LGM factors. For a computed interaction model, Contact × ASP × Age did not predict growth in noncompliance over and above the significant effect of Contact × ASP. However, the effect for Contact × ASP × Age on the average level intercept factor obtained a significant standardized beta ($\beta = -.159$, $t = -2.09$, $p < .04$). This suggested that the impact of antisocial fathers' contact was lessened for older children relative to younger children in predicting individual differences in average levels of noncompliance. This is consistent with the notion that younger children are both more noncompliant and more vulnerable to the effects of divorce. However, inspection of the explained variance showed that this three-way interaction contributed an additional 1% explained variance to the model predicting average levels. Using latent variable interaction terms, the two-way interaction effect of Contact × ASP was replicated in predicting the intercept ($t = 2.065$) and the growth factor ($t = 2.694$). However, the effect of Contact × ASP × Age on the intercepts term was marginal using the latent variable interaction approach ($t = -1.78$, $p < .08$).

In summary, the findings showed that the effect of fathers' ASP moderating his monthly contact was robust in predicting intercepts and growth rates in noncompliance for boys and girls and across child age. Results for the average levels of noncompliance were somewhat stronger for younger children relative to older, but this effect was not very meaningful in that one test produced an additional 1% of explained variance and the latent variable interaction test was marginal. To illustrate the main finding, the Contact × ASP effect was plotted using the estimated growth slopes from Figure 3 for 1 *SD* above the mean of ASP and 1 *SD* below the mean (Figure 4).

Discussion

This study investigated the impact of father involvement indexed as a continuum of monthly contact with his child following divorce and attempted to add to the previous literature by focusing on fathers of young children who are more vulnerable to behavioral effects of divorce. By comparison, a majority of prior research has focused less on fathers' actual parenting behaviors and more so on father contact only. The present study employed multiple-method assessment

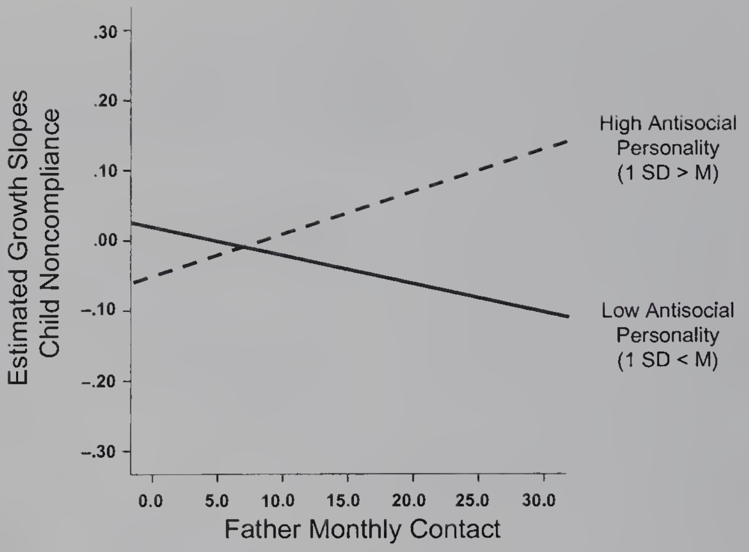


Figure 4. Plot of estimated linear growth slopes for effect of father antisocial personality as a moderator of monthly contact. *Note.* Slopes are plotted for 1 *SD* above the mean of antisocial personality and 1 *SD* below the mean.

including direct observation of fathering and child noncompliance. Furthermore, extant research has focused mainly on the period of adolescence.

The theoretical model employed the coercion model of child development and the generative theories of father involvement. Fathers that have regular contact with their children can have a positive developmental impact. However, fathers' ASP was expected to moderate monthly contact, and ASP was expected to be associated with higher coercive parenting.

The unconditional growth models showed significant individual differences in children's noncompliance and individual variation in trajectories of increases and decreases. Supporting the hypotheses, latent growth prediction models showed that monthly contact was negatively associated with children's growth in noncompliance at average levels of antisociality (ASP); however, father ASP significantly moderated the effect of monthly contact. As shown in Figure 4, increased contact of low-antisocial fathers was associated with decreases in noncompliance over 18 months and contact with high-antisocial fathers was associated with increases. Exploratory models showed that this effect was robust for younger and older children and for boys and girls.

Partially supporting the hypotheses, fathers' ASP predicted higher levels of coercive parenting practices, and coercive parenting predicted average levels of noncompliance but not growth rates. Conversely, it should be noted that latent variable models also imply that effective prosocial parenting practices were associated with lower levels of noncompliance, underscoring the importance of prosocial parenting skills. From a clinical perspective, it is important to note that the coercion model is complementary and not antithetical to a generative fathering approach. Preventive intervention approaches based on the coercion model emphasize positive parenting as the cornerstone to reducing coercive parenting, including positive reinforcement, skill encouragement, and emotion regulation (Patterson, 2005).

Contrary to expectations, parenting practices did not account for the interaction effect of Fathers' Contact × ASP. Although this study was not a genetically informed design, the pattern of findings replicated what Jaffee et al. (2003) described as the ''double whammy'' of environmental and genetic influence of antisocial fathers. Given that parenting did not account for the ASP interaction, this suggested a potential genetic influence on children's problem behaviors. At the same time, controlling for ASP, parenting practices explained an additional 14% of the variance in levels of noncompliance, underscoring the importance of fathers' parenting practices. Parenting did not predict variance in growth trajectories.

Limitations

A key limitation existed regarding the prediction of parenting on change in noncompliance. Predivorce levels of parenting and noncompliance were not collected. It is likely parenting effects on growth in noncompliance were more influential during earlier socialization prior to divorce. The fact that fathers' contact conditioned by ASP was predictive of increases in noncompliance also suggested that other potential mechanisms in addition to genetic influence may account for the effects of father's monthly contact effect, such as the child's experience of socioemotional closeness with fathers or conversely subjective distress associated with divorce-related stressors between parents (Wolchik, Tein, Sandler, & Doyle, 2002).

It is important to note that the present study was based on public court records with a divorce decree from a county sampling frame including rural and urban areas. There are known limitations to court records based studies involving participation of families who separate and do not file for divorce and inferences based on families in which the marital dissolution and disruptions to social interactional may have occurred long before filing. As noted earlier, the present study was limited because it was not a prospective study of divorce and did not have measures of children's predivorce functioning, parenting, or mothers' parenting, all of which are probably influencing the average levels of children's problem behaviors and growth. For shared- and no-custody fathers, mothers' parenting is likely a key predictor children's noncompliance. Furthermore, although recent studies have shown independent effects of nonresidential fathers net of mothers, it is likely that effects of fathers' antisociality may share variance with mother's antisociality influences due to assertive mating.

Another limitation was the measure of monthly contact. A more refined measure of father contact would be time-use diary approaches for measuring father involvement (e.g., Hofferth, Stueve, Pleck, Bianchi, & Sayer, 2002). Recently, however, Wical and Doherty (2005) have shown that fathers are reliable reporters of their contact. Nonetheless, given the limitations, the present findings replicated associations reported for adolescence and included direct observation of fathers' parenting and child adjustment. The findings underscored the impact of residential and nonresidential fathers on children's behaviors following divorce and potential mechanisms of preventive intervention.

Implications

Prevention programs may be designed to be universal, selective, or indicated programs to address child adjustment problems in the hopes of delaying or preventing onset of problem behaviors. As opposed to universal programs for an entire population, selective prevention programs target subsets of the population that are at risk by a common membership, for example, children of divorced parents. As opposed to treatment, selective prevention targets individuals in a group (divorced fathers) regardless of individual risk or without prior screening. Indicated prevention focuses on individuals from a risk group that are showing early signs of problems (e.g., deficient parenting or onset of conduct disorder in children). Currently, many individual and community-based programs for promoting effective fathering and preventing child adjustment problems exist due largely in response to the Department of Health and Human Services Fatherhood Initiative (National Center on Fathers and Families, 2006). The majority of fatherhood programs, however, have limited theoretical or evidence-based foundations (McBride & Lutz, 2004).

Exceptions are programs focusing on father involvement and the reduction of coparenting conflict, the Dads for Life (DFL) program (Braver, Griffin, & Cookston, 2005) and the Strengthening Father Involvement (SFI) program (Cowan, Cowan, Pruett, & Pruett, 2007). Like the DFL, the SFI has demonstrated reductions in coparenting conflict over time and more recently has shown benefits to children's problem behaviors (Cowan, Cowan, Pruett, Pruett, & Wong, 2009). From a parent training perspective, the Oregon Model of Parent Training (PMTO) is another exception (Patterson, 2005). To date, the Forgatch et al. (2005) intervention was only one of two theory- and evidence-based parent training programs involving stepfathers. One critical finding was that the intervention produced the largest effect size on improved parenting behaviors for stepfathers when compared to effects for mothers (DeGarmo & Forgatch, 2007). It is likely that a selective preventive intervention program for divorced fathers would benefit children's behavioral adjustment over time.

In addition to direct work with fathers or couples, approaches involving direct work with children incorporating father involvement may benefit divorced families. For example, parent–child interaction therapy effectively reduces children's problem behaviors. Recent efforts have shown benefits for maintenance of effects involving components of father involvement (Bagner & Eyberg, 2003).

Generative fathering perspectives may be helpful in delivering prevention programs to antisocial fathers. Dishion et al. (2004) argued that it is likely that deviant fathers are underrepresented in prevention research, and currently little research exists on how antisocial fathers respond to parenting interventions and may, therefore, show larger effects. Examples from clinical treatment work with maltreating and abusive fathers have shown that fathers are more responsive to interventions pointing out the welfare and needs of the children as a focal point for addressing harmful interactions with spouses and children (Rosenberg & Wilcox, 2006; Scott & Crooks, 2004). A coupling of the social interactional–social learning intervention model with generative fathering perspectives might show great promise. Although the coercion model focuses on a deficit of skilled parenting, the model is not antithetical to generative fathering perspectives in its theoretical approach. For example, PMTO (Patterson, 2005) focuses on positive reinforcement, encouragement, and positive involvement as cornerstones for reductions in coercive parenting and child noncompliance. Because of fathers' desire to stay involved with their children, effective programs may benefit from father-oriented components increasing men's awareness of child-centered needs and parenting (Brotherson et al., 2005; Parke & Brott, 1999).

Finally, given the main effect of parenting practices, it is argued the present findings do not imply that only antisocial fathers would benefit from parenting intervention or that all divorced fathers are in need of parent training. The present findings suggest that both promoting more effective parenting skills would benefit a sample of divorced fathers and a focus on antisociality would be important. Prevention programs are designed to address populations with a range of parenting skills as opposed to screening and targeting fathers with parenting deficiencies. Evidence-based programs have benefited single mothers and their children using community and court-mandated samples and selected and indicated samples, all of which comprised mothers ranging in parenting skill and antisociality. Programmatic intervention with divorced fathers should also prove independently beneficial for divorced families members.

References

Amato, P. R., & Gilbreth, J. G. (1999). Nonresident fathers and children's well-being: A meta-analysis. *Journal of Marriage and the Family*, *61*, 557–573.

Bagner, D. M., & Eyberg, S. M. (2003). Father involvement in parent training: When does it matter? *Journal of Clinical Child and Adolescent Psychology*, *32*, 599–605.

Biesanz, J. C., Deeb-Sossa, N., Papadakis, A. A., Bollen, K. A., & Curran, P. J. (2004). The role of coding time in estimating and interpreting growth curve models. *Psychological Methods*, *9*, 30–52.

Braver, S. L., Griffin, W. A., & Cookston, J. T. (2005). Prevention programs for divorced nonresident fathers. *Family Court Review*, *43*, 81–96.

Braver, S. L., Wolchick, S. A., Sandler, I. N., Sheets, V. L., Fogas, B., & Bay, R. C. (1993). A longitudinal study of nonresidential parents: Parents without children. *Journal of Family Psychology*, *7*, 9–23.

Brotherson, S. E., Dollahite, D. C., & Hawkins, A. J. (2005). Generative fathering and the dynamics of connection between fathers and their children. *Fathering*, *3*, 1–28.

Brown, H. C., Wang, W., Kellam, S. G., Muthén, B. O., Petras, H., Toyinbo, P., et al. (2008). Methods for testing theory and evaluating impact in randomized field trials: Intent-to-treat analyses for integrating the perspectives of person, place, and time. *Drug and Alcohol Dependence*, *95*(Suppl. 1), S74–S104.

Capaldi, D. M., & Patterson, G. R. (1991). Relation of parental transitions to boys' adjustment problems: I. A linear hypothesis. II. Mothers at risk for transitions and unskilled parenting. *Developmental Psychology*, *27*, 489–504.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Coley, R. L., & Medeiros, B. L. (2007). Reciprocal longitudinal relations between nonresident father involvement and adolescent delinquency. *Child Development, 78*, 132–147.

Costa, P. T. Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.

Costa, P. T., Jr., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the revised NEO Personality Inventory. *Journal of Personality Assessment, 64*, 21–50.

Cowan, C. P., Cowan, P. A., Pruett, M. K., & Pruett, K. (2007). An approach to preventing coparenting conflict and divorce in low-income families: Strengthening couple relationships and fostering fathers' involvement. *Family Process, 46*, 109–121.

Cowan, P. A., Cowan, C. P., Pruett, M. K., Pruett, K., & Wong, J. J. (in press). Promoting fathers' engagement with children: Preventive interventions for low-income families. *Journal of Marriage and the Family*, 663–679.

DeGarmo, D. S., & Forgatch, M. S. (2007). Efficacy of parent training for stepfathers: From playful spectator and polite stranger to effective stepfathering. *Parenting: Science and Practice, 7*, 1–25.

DeGarmo, D. S., Patras, J., & Eap, S. (2008). Social Support for divorced fathers' parenting: Testing a stress buffering model. *Family Relations, 57*, 35–48.

Dishion, T. J., Owen, L. D., & Bullock, B. M. (2004). Like father, like son: Toward a developmental model for the transmission of male deviance across generations. *European Journal of Developmental Psychology, 1*, 105–126.

Dunn, J., Cheng, H., O'Connor, T. G., & Bridges, L. (2004). Children's perspectives on their relationships with their nonresident fathers: Influences, outcomes and implications. *Journal of Child Psychology and Psychiatry, 45*, 553–566.

Forgatch, M. S., DeGarmo, D. S., & Beldavs, Z. G. (2005). An efficacious theory-based intervention for stepfamilies. *Behavior Therapy, 36*, 357–365.

Hawkins, A. J., & Dollahite, D. C. (Eds.). (1997). Beyond the role-inadequacy perspective of fathering. In *Generative fathering: beyond deficit perspectives* (pp. 3–16). Thousand Oaks, CA: Sage.

Hetherington, E. M., & Clingempeel, W. G. (1992). Coping with marital transitions. *Monographs of the Society for Research in Child Development, 57*(2–3, Serial No. 227).

Hofferth, S. L., Stueve, J. L., Pleck, J., Bianchi, S., & Sayer, L. (2002). The demography of fathers: What fathers do. In C. S. Tamis-LeMonda & N. Cabrera (Eds.), *Handbook of father involvement: Multidisciplinary perspectives* (pp. 63–90). Mahwah, NJ: Erlbaum.

Jaffee, S. R., Moffitt, T. E., Caspi, A., & Taylor, A. (2003). Life with (or without) father: The benefits of living with two biological parents depend on the father's antisocial behavior. *Child Development, 74*, 109–126.

King, V., & Sobolewski, J. M. (2006). Nonresident fathers' contributions to adolescent well-being. *Journal of Marriage and Family, 68*, 537–557.

Martinez, C. R., Jr., & Forgatch, M. S. (2001). Preventing problems with boys' noncompliance: Effects of a parent training intervention for divorcing mothers. *Journal of Consulting and Clinical Psychology, 69*, 416–428.

McBride, B. A., & Lutz, M. M. (2004). Intervention: Changing the nature and extent of father involvement. In M. E. Lamb (Ed.), *The role of the father in child development* (4th ed., pp. 446–475). Hoboken, NJ: Wiley.

Muthén, L. K., & Muthén, B. O. (2007). *MPlus: Statistical analysis with latent variables user's guide* (5th ed.). Los Angeles: StatModel.

National Center on Fathers and Families. (2006). Survey of father and family related programs. Retrieved January 19, 2010, from http://www.ncoff.gse.upenn.edu/programsdb/progsrch.htm

Parke, R. D., & Brott, A. A. (1999). *Throwaway dads: The myths and barriers that keep men from being the fathers they want to be*. New York: Houghton Mifflin.

Parke, R. D., Dennis, J., Flyr, M. L., Morris, K. L., Killian, C., McDowell, D. J., et al. (2004). Fathering and children's peer relationships. In M. E. Lamb (Ed.), *The role of the father in child development* (pp. 307–340). Hoboken, NJ: Wiley.

Pasley, K., & Braver, S. L. (2004). Measuring father involvement in divorced, nonresident fathers. In R. D. Day & M. E. Lamb (Eds.), *Conceptualizing and measuring father involvement* (pp. 217–240). Mahwah, NJ: Erlbaum.

Pasley, K., & Minton, C. (1997). Generative fathering after divorce and remarriage: Beyond the "disappearing dad." In A. J. Hawkins & D. C. Dollahite (Eds.), *Generative fathering: Beyond deficit models* (pp. 118–133). Thousand Oaks, CA: Sage.

Patterson, G. R. (1982). *Coercive family process*. Eugene, OR: Castalia.

Patterson, G. R. (2005). The next generation of PMTO models. *Behavior Therapist, 28*, 25–32.

Patterson, G. R., & Dishion, T. J. (1988). Multilevel family process models: Traits, interactions, and relationships. In R. A. Hinde & J. Stevenson-Hinde (Eds.), *Relationships within families: Mutual influences* (pp. 283–310). Oxford, UK: Clarendon Press.

Patterson, G. R., & Yoerger, K. (1999). Intraindavidual growth in covert antisocial behaviour: A necessary precursor to chronic juvenile and adult arrests? *Criminal Behavior and Mental Health, 9*, 24–38.

Reid, J. B., Patterson, G. R., & Snyder, J. (Eds.). (2002). *Antisocial behavior in children and adolescents: A developmental analysis and model for intervention*. Washington, DC: American Psychological Association.

Rosenberg, J., & Wilcox, W. B. (2006). *The importance of fathers in the healthy development of children: Child abuse and neglect user manual series*. Washington, DC: U.S.

Department of Health and Human Services, Children's Bureau.

Ronald, L. Simons and Associates. (1996). *Understanding differences between divorced and intact families.* Thousand Oaks, CA: Sage.

Scott, K. L., & Crooks, C. V. (2004). Effecting change in maltreating fathers: Critical principles for intervention planning. *Clinical Psychology: Science and Practice, 11,* 95–111.

Simons, R. L., Beaman, J., Conger, R. D., & Chao, W. (1993). Stress, support, and antisocial behavior trait as determinants of emotional well-being and parenting practices among single mothers. *Journal of Marriage and the Family, 55,* 385–398.

Simons, R. L., Whitbeck, L. B., Beaman, J., & Conger, R. D. (1994). The impact of mothers' parenting, involvement by nonresidential fathers, and parental conflict on the adjustment of adolescent children. *Journal of Marriage and the Family, 56,* 356–374.

Stephens, L. (1996). Will Jonny see daddy this week? *Journal of Family Issues, 17,* 466–494.

Stubbs, J., Crosby, L., Forgatch, M. S., & Capaldi, D. M. (1998). *Family and Peer Process Code: A synthesis of three Oregon Social Learning Center behavior codes* (Training manual). Eugene, OR: Oregon Social Learning Center.

Swanson, S. C., Templer, D. I., Streiner, D. L., Reynolds, R. M., & Miller, H. R. (1995). Development of a three-scale MMPI: The MMPI-TRI. *Journal of Clinical Psychology, 51,* 361–374.

Vuchinich, S., Hetherington, E. M., Vuchinich, R. A., & Clingempeel, W. G. (1991). Parent-child interaction and gender differences in early adolescents' adaptation to stepfamilies. *Developmental Psychology, 27,* 618–626.

Wical, K. A., & Doherty, W. J. (2005). How reliable are fathers' reports of involvement with their children? *Fathering, 3,* 81–91.

Wolchik, S. A., Tein, J.-Y., Sandler, I. N., & Doyle, K. W. (2002). Fear of abandonment as a mediator of the relations between divorce stressors and mother-child relationship quality and children's adjustment problems. *Journal of Abnormal Child Psychology, 30,* 401–418.

Wothke, W. (2000). Longitudinal and multi-group modeling with missing data. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multiple group data: Practical issues, and specific examples* (pp. 269–281). Mahwah, NJ: Erlbaum.

Child Development, March/April 2010, Volume 81, Number 2, Pages 517–527

The Tuning of Human Neonates' Preference for Speech

Athena Vouloumanos
New York University

Marc D. Hauser
Harvard University

Janet F. Werker
University of British Columbia

Alia Martin
New York University

Human neonates prefer listening to speech compared to many nonspeech sounds, suggesting that humans are born with a bias for speech. However, neonates' preference may derive from properties of speech that are not unique but instead are shared with the vocalizations of other species. To test this, thirty neonates and sixteen 3-month-olds were presented with nonsense speech and rhesus monkey vocalizations. Neonates showed no preference for speech over rhesus vocalizations but showed a preference for both these sounds over synthetic sounds. In contrast, 3-month-olds preferred speech to rhesus vocalizations. Neonates' initial biases minimally include speech and monkey vocalizations. These listening preferences are sharpened over 3 months, yielding a species-specific preference for speech, paralleling findings on infant face perception.

How do initial perceptual biases and learning from specific experience interact to prepare infants to adapt optimally to their world? Across perceptual domains, infants use early biases to organize the world. For example, young infants are able to discriminate non-native speech sounds that are unfamiliar to them (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Werker & Tees, 1984) and identify individual faces of species they have never encountered before (Pascalis, de Haan, & Nelson, 2002). By the end of the 1st year of life, these biases become attuned to sounds and visual information in the infant's environment, such that infants no longer readily discriminate non-native speech sounds (Kuhl et al., 1992; Werker & Tees, 1984) nor do they identify unfamiliar species' faces (Pascalis et al., 2002). Initial biases, and rapid attunement, may constitute conserved and fundamental principles that underwrite the development of perceptual systems across the animal kingdom. In this article, we explore the ontogeny of the specificity of the neonatal preference for human speech.

Many species are endowed with an innate preference to attend to the visual and auditory cues of their conspecifics (Gottlieb, 1997; Johnson, Bolhuis, & Horn, 1985; Marler & Peters, 1977). For example, precocial birds preferentially approach the calls of conspecifics, whereas the young fledglings of song learning birds will selectively acquire their species-specific song if they are exposed to it during a sensitive period (Gottlieb, 1997; Marler & Peters, 1977). A common perspective in studies of language acquisition is that humans are similarly endowed with a preference for conspecific vocalizations and, more specifically, that speech is special for humans (Liberman & Whalen, 2000; Pinker & Jackendoff, 2005), perceived in unique ways from the earliest stages of development (Vouloumanos & Werker, 2007). Support for this hypothesis comes from studies showing that neonates exhibit a preference for speech over filtered (Spence & DeCasper, 1987) and reversed speech (Peña et al., 2003), as well as synthesized signals such as complex synthetic sounds (Vouloumanos & Werker, 2007) and white noise

---

Funding was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC), the Centre for Research in Language, Mind and Brain, and the Fonds québécois de recherche sur la société et la culture to A. V.; by the James S. McDonnell Foundation and 21st Century grants to M. D. H.; by NSERC, the Canada Research Chair Program, and the James S. McDonnell Foundation to J. F. W., and by an NSERC Undergraduate Summer Research Award to A. M. M. D. H.'s research was also funded by gifts from J. Epstein and S. Shuman. We thank Erin Moon, Marisa Cruickshank, and Vivian Pan for help with data collection. For comments on earlier versions of the manuscript, we thank Steven Pinker and especially Gary Marcus. We are especially grateful to participating parents and infants.

Correspondence concerning this article should be addressed to Athena Vouloumanos, Department of Psychology, New York University, 6 Washington Place, New York, NY 10003. Electronic mail may be sent to athena.vouloumanos@nyu.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8102-0009

(Butterfield & Siperstein, 1970). Such a preference for speech suggests that infants are prepared to attend to auditory stimuli in their environment that are relevant for language, thus facilitating language acquisition (Kuhl, Coffey-Corina, Padden, & Dawson, 2005).

Although previous studies support the hypothesis that speech is a privileged sound for infants, they leave open the possibility that neonates' listening preference is based on properties of speech that are not unique to speech but instead are shared with the vocalizations of related species. Speech and nonhuman primate vocalizations share some acoustic properties and are coded similarly in the human auditory system (Altmann, Doehrmann, & Kaiser, 2007; Smith & Lewicki, 2006). We selected rhesus monkey calls for comparison because, although acoustically distinct from human speech, they are similarly spectrally rich. In particular, rhesus have a broad vocal repertoire spanning a wide range of activities, including calls produced during the discovery of food sources (e.g., coos, warbles) and affiliative social behavior (e.g., girneys; Hauser, 2000). Like human speech, some rhesus monkey calls are produced by oscillations of the larynx (phonation) and filtered by the supralaryngeal vocal tract, which dampens or amplifies different frequencies (Fitch & Hauser, 1995; Hauser, Evans, & Marler, 1993; Rendall, Owren, & Rodman, 1998; see Figure 1). This filtering of frequencies generated by the laryngeal oscillations produces discrete formants, or concentrations of energy, whose distribution and fluctuations across time generate a wide range of speech sounds. As a result of constraints on the movements of rhesus articulators, however, the "phonetic" repertoire of rhesus monkeys is limited by comparison with that of humans (Belin, 2006; Fitch, 2000) and the frequency spectrum of calls is broader and higher so that, perceptually, rhesus calls present a broader distribution of spectral energy and a higher pitch than human speech (Fitch, 1997; Owren & Rendall, 2003). Although some rhesus calls have an acoustic resemblance to human affective vocalizations, adult human listeners do not perceive rhesus calls as being generated by humans (see ratings in the Method section that follows), nor do they explicitly distinguish these calls based on their affective valence (Belin et al., 2008). Testing a preference for speech against the vocalizations of a closely related primate provides a strong test for an early specialization for speech in humans.

Using a contingency procedure in which infant participants could control the presentation of different sounds, we tested neonates on their preference for speech compared to the vocalizations of another primate—the rhesus macaque (*Macaca mulatta*). There are three possible outcomes. First, if neonates prefer speech to rhesus calls, then this supports earlier reports that speech is privileged for human neonates, distinct from the vocalizations of another related species. The neonatal preference for speech would then be based on properties that are not shared with rhesus calls. Second, if neonates prefer rhesus calls to speech, then the task is most likely picking up on a novelty preference. Although this would be a surprising outcome as previous studies showed a neonatal preference for speech over many other novel sounds, previous studies never evaluated the preference for human speech against the calls of a related species. Finally, if neonates show equal interest in speech and rhesus calls, with no preference for one sound over the other, then this suggests that the initial bias for speech may derive from properties shared by speech and nonhuman primate vocalizations. In this case, to ensure that the lack of preference was not simply the result of a methodological failure, it would be important to show—as in earlier work on the neonatal speech preference—that neonates also prefer rhesus vocalizations to other acoustic signals. Finally, we

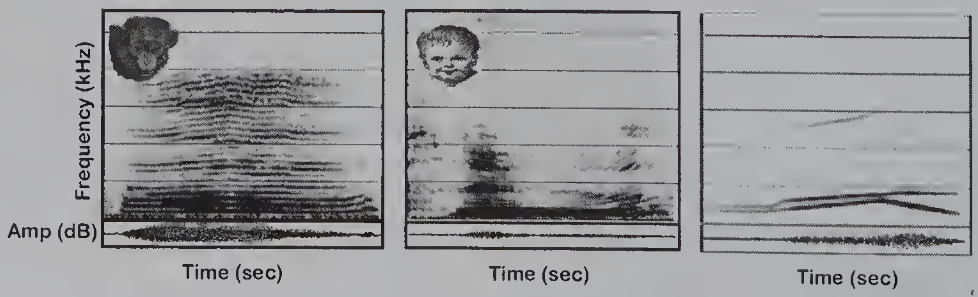


Figure 1. Spectrogram (top) and waveform (bottom) of a rhesus monkey vocalization (coo; on the left), a human speech token ("ploo"; in the middle), and a synthetic nonspeech sound (on the right).
Note. Horizontal lines denote frequency increments of 2000 Hz.

examine the listening preferences of 3-month-old infants.

Experiment 1

We tested human neonates on their preference for speech compared to rhesus calls using an infant-controlled high-amplitude (HA) sucking method (e.g., Butterfield & Siperstein, 1970; DeCasper & Fifer, 1980; Moon, Cooper, & Fifer, 1993; Vouloumanos & Werker, 2007). In this procedure, neonates sucked on a sterilized non-nutritive pacifier to control the presentation of sounds. We presented nonsense speech using English-like speech syllables and rhesus monkey calls in alternating trials, and measured the number of neonates' HA sucks for each of the two types of vocalizations.

Method

Participants

Thirty full-term (minimum of 37.5 weeks gestational age) human neonates (M = 47 hr, range = 18–120 hr; 12 females and 18 males) completed the study. An additional 44 neonates were excluded for noncompliant behavior, such as fussing or falling asleep (24), parental interference (1), or failing to meet established experimental criteria (Floccia, Nazzi, & Bertoncini, 2000; Vouloumanos & Werker, 2007) of a minimum of 10 HA sucks in the first 4 min and at least 1 suck in every subsequent experimental minute (19). Neonates were tested at a birthing Hospital in Vancouver, British Columbia. Reflecting the diversity of the local community, the languages spoken in neonates' home environment varied, with the percentage of English spoken between 0% and 100%. We did not collect additional demographic information about family ethnicity. Parents gave informed consent on behalf of their infants. All procedures were approved by the university's Research Ethics Board.

Stimuli

In selecting specific monkey and human stimuli, it was important to control for factors that in and of themselves might bias infants toward one or the other group of sounds. As neonates listen less to angry and sad voices than they do to happy or neutral voices (Aldridge, 1994, as cited in Walker-Andrews, 1997), we excluded rhesus calls produced during aggressive behaviors (e.g., screams, barks, and copulatory screams). Instead, we selected calls produced during positive social contexts such as affiliative interactions (girneys, coos) and the discovery of food (coos, warbles; Hauser, 2000) considered to be of positive affect in previous studies (Belin et al., 2008; Nicastro & Owren, 2003).

Speech. Ten nonsense speech sounds were produced in infant-directed speech by three native English female speakers (Speaker 1: ''ploo,'' ''keev,'' ''yut,'' ''boola,'' ''nahod,'' ''kraw''; Speaker 2: ''trom,'' ''fi''; Speaker 3: ''dup,'' ''makalak'') with average pitch of 283 Hz, and average duration of 468 ms.

Monkey calls. Ten rhesus monkey calls were recorded on Cayo Santiago, Puerto Rico, from three female rhesus monkeys (Caller 1: 3 coos [social or food related], 3 girneys [affiliative]; Caller 2: 1 coo, 1 warble [food related]; Caller 3: 1 coo, 1 girney) with average pitch of 346 Hz and average duration of 531 ms. Calls were recorded with a highly directional shotgun microphone at a distance of 1–3 m. The calls were selected such that there was a high signal-to-noise ratio, only one concurrent vocalization, and no interference from other biotic sounds. Stimulus duration did not differ, $t(18) < 1$, *ns*, but rhesus calls naturally exhibit a slightly higher pitch, $t(17) = 2.3$, $p < .05$. (For description of acoustic morphology and context, see Hauser, 1992, 1996; Hauser & Fowler, 1992; Hauser & Marler, 1993.)

Adult ratings of vocalizations. Thirty undergraduates rated the 20 primate vocalizations described on a 5-point scale for how likely they were to have been produced by a human. The 10 monkey calls were rated as significantly less human-like than the 10 human nonsense speech ($M = 2.97$, $SE = 0.924$ for monkey vocalizations; $M = 4.63$, $SE = 0.122$ for human nonsense speech), $t(18) = 5.61$, $p < .001$.

Apparatus and Procedure

Neonatal listening preferences were quantified by means of a HA sucking procedure (Eimas, Siqueland, Jusczyk, & Vigorito, 1971) in which the presentation of different sounds was contingent on the newborn delivering an HA suck (Vouloumanos & Werker, 2007). Approximately 2 hr after feeding, neonates were brought to a sound-attenuated room and placed supine in a bassinette. During the experiment, a parent was seated behind and out of reach of the bassinette. An experimenter wearing headphones playing music presented neonates with a sterilized pacifier coupled to a Gould P23 pressure transducer, which was connected to an IBM-PC-compatible computer. After presenting the pacifier to the neonate, the experimenter only interacted with the newborn to reinsert the pacifier if it had been released. If the pacifier could not be reinserted immediately, the neonate's data were discarded from the sample. The computer registered air pressure changes caused by infant sucks.

Each individual infant's sucking amplitude range was measured during a silent baseline minute. The mean baseline HA sucking rate was 33.7 HA sucks/min (*SE* = 2.93). Following the calculation of their baseline sucking amplitude, neonates were presented with a sound stimulus whenever they applied suction in the upper 80% of their sucking amplitude range. The stimulus onset asynchrony (from the start of one stimulus to that of the next) was set at 1500 ms to accommodate the longest sound duration. The presentation rate was thus one stimulus per 1.5 s. All sounds were presented at 70 dB (±4 dB). The presentation of speech and rhesus calls was alternated every minute, with order counterbalanced across participants, and the 8 min postbaseline submitted to analysis. The HA sucking software implements the minute-by-minute change independently of the baby's sucking behavior as in many HA sucking experiments (Bertoncini, Bijeljac-Babic, Blumstein, & Mehler, 1987; Bertoncini, Bijeljac-Babic, Jusczyk, Kennedy, & Mehler, 1988; Floccia et al., 2000; Mehler et al., 1988; Nazzi, Floccia, & Bertoncini, 1998). To ensure that newborns heard sounds during every minute, we excluded newborns that produced fewer than 10 sucks in the first four experimental minutes or 0 sucks during any of the 8 min. Within each minute, tokens were presented randomly with replacement allowing repetitions of a given token. Earlier applications of this procedure have been successful in showing that newborns prefer speech over complex synthetic sounds (Vouloumanos & Werker, 2007).

Results and Discussion

As there were no significant effects of infant sex or sound order on listening preferences ($F < 1$, *ns*), we collapsed across these factors for all analyses. As sampling over blocks can yield more stable measurements and has successfully revealed neonatal preferences in previous studies (Vouloumanos & Werker, 2007), we analyzed HA sucks in two observation blocks (first 4-min block vs. second 4-min block) in a repeated measures analysis of variance (ANOVA). Results indicated no main effects and no interactions (all *ns*). Neonates sucked equivalently for speech ($M = 33.3$ HA sucks, $SE = 2.4$) and for rhesus calls ($M = 33.8$ HA sucks, $SE = 2.7$), $F(1, 29) < 1$, *ns*. An additional ANOVA analyzing preference for speech and monkey calls for each experimental minute also revealed no main effects and no interactions (all *ns*). Next, we examined whether neonates' HA sucking changed over the two observation blocks for speech or for rhesus calls and found no statistically significant effect (using Bonferroni corrections for multiple comparisons) for either speech, $t(29) = 2.12$, *ns*, or rhesus calls $t(29) = 1.02$, *ns* (see Table 1). Thirteen of 30 neonates exhibited a preference for speech (see Figure 2). To examine whether the amount of English exposure in the neonates' home environment had an effect on neonates' listening preferences, we ran an ANOVA with sound type (speech, rhesus) and percentage of English exposure (as estimated by parental report at the time of consent) as a covariate and found no main effect of English exposure, $F(1, 28) < 1$, *ns*, and no modulatory effect of English exposure on neonates' listening preference for speech and rhesus calls, $F(1,28) = 1.19$, *ns*.

Table 1

High-Amplitude (HA) Sucking Across the Three Studies: (a) Speech Versus Rhesus Monkey Calls, (b) Rhesus Monkey Calls Versus Synthetic Sounds, and (c) Speech Versus Synthetic Sounds (Reported in Vouloumanos & Werker, 2007)

Sound comparison	*n*	Sound	HA Sucking Block 1	HA Sucking Block 2	Relative change in HA sucking
Speech versus monkey calls	30	Speech	71.20	61.73	−4.4
		Monkey calls	69.90	64.83	
Monkey calls versus synthetic sounds	16	Monkey calls	68.75	60.50	12.3*
		Synthetic sounds	74.50	54.00*	
Speech versus synthetic sounds[a]	22	Speech	71.91	73.00	14.5*
		Synthetic sounds	76.91	63.45*	

[a] Previously reported in Vouloumanos and Werker (2007).

Note. Neonates' HA sucking decreases significantly across blocks for the synthetic sounds; no significant decrease across blocks is observed for either speech or rhesus monkey calls. The relative change in HA sucking in the speech versus rhesus call comparison is significantly different both from the monkey call versus synthetic sound comparison, and from the speech versus synthetic sound comparison. The relative change in HA sucking for monkey over synthetic sounds is equal to that for speech over synthetic sounds. See the Results section for more details.

$^{*}p < .05$.

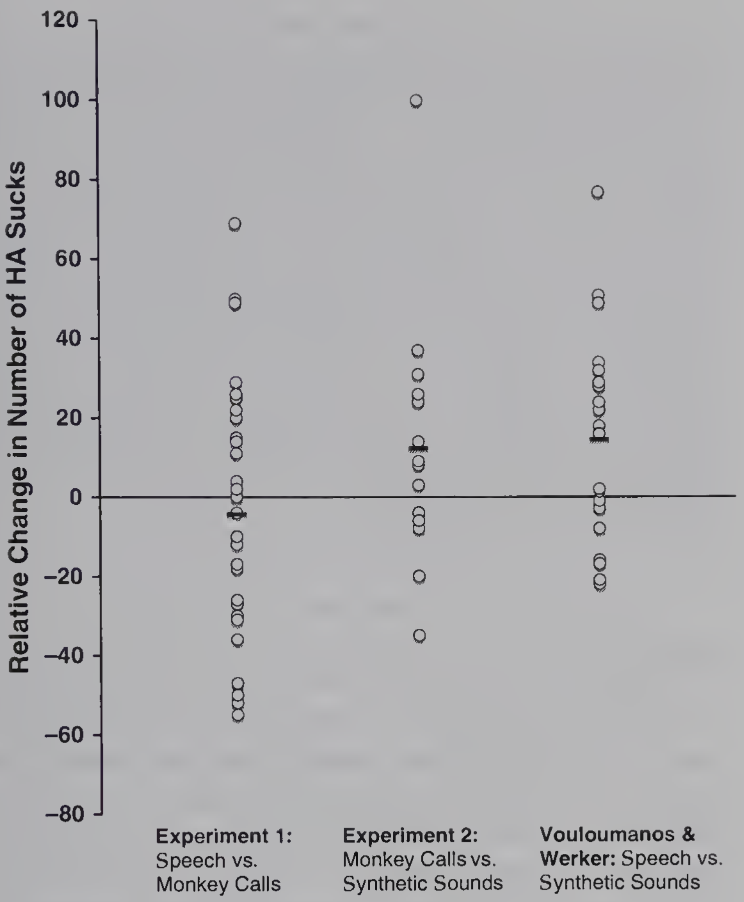


Figure 2. Scatterplot of individual neonates' relative change in the number of high-amplitude sucks for sound types across the three experiments.

Note. For Experiment 1, we plot as positive the relative change for speech over monkey calls; for Experiment 2, we plot as positive the relative change for monkey calls over synthetic sounds; and for Vouloumanos and Werker (2007), we plot as positive the relative change for speech over synthetic sounds.

Although earlier work showed that neonates preferred speech to artificial sounds and manipulated speech sounds (Butterfield & Siperstein, 1970; Peña et al., 2003; Spence & DeCasper, 1987; Vouloumanos & Werker, 2007), this study demonstrates that neonates have no preference for speech over rhesus calls, responding equivalently to both.

Experiment 2

To ensure that the equivalent interest in speech and rhesus calls reflects a legitimate preference for listening to these two types of vocalizations, as opposed to a null result, we ran an additional condition. In a previous study, neonates preferred speech to complex synthetic sounds (Vouloumanos & Werker, 2007). These complex synthetic sounds were composed of four sinusoidal waves that emulated the frequency changes of speech and thus contained a similar pitch contour and some of its spectral features (Vouloumanos & Werker, 2004, 2007); in some contexts, adult listeners can even perceive these types of complex synthetic sounds as speech-like (Remez, Rubin, Pisoni, & Carrell, 1981). In this previous study, neonates manifested a preference for speech by adjusting their sucking rate over the course of the experiment so as to listen to speech more than to the complex synthetic sounds. Over the two observation blocks, neonates in this previous study maintained their HA sucking rate for speech, $t(21) = 0.315$, *ns*, but significantly decreased their sucking for the complex synthetic sounds, $t(21) = 2.92$, $p < .02$ (Vouloumanos & Werker, 2007). To test whether human neonates show a similar sustained interest in listening to rhesus monkey calls, we presented rhesus monkey calls and complex synthetic sounds to 16 neonates. If both speech and rhesus calls are preferred vocalizations for human neonates, then neonates' HA sucking in this experiment should show differential arousal to rhesus monkey calls over complex synthetic sounds.

Method

Participants

Sixteen full-term (minimum of 37.5 weeks gestational age) human neonates (M = 46 h, range = 20–97 hr; 7 females and 9 males) completed the study. An additional 21 neonates were excluded for fussing or falling asleep (10), parental interference (1), or not meeting the established experimental criteria (10). Neonates were tested at a birthing hospital in Vancouver, British Columbia. As in Experiment 1, the percentage of English spoken in neonates' home environment varied between 0% and 100%, and additional ethnographic information was not collected. Parents gave informed consent on behalf of their infants. All procedures were approved by the university's Research Ethics Board.

Stimuli

Monkey calls. These were identical to those used in Experiment 1. As there were four different complex synthetic tokens in total, we randomly selected a different subset of 4 monkey tokens from the full set of 10 tokens for presentation for each neonate.

Synthetic sounds. Synthetic sounds consisted of time-varying sinusoidal waves tracking the main regions of significant energy, namely, the fundamental frequency and the first three formants of a nonsense speech token. By tracking changes across time for the peak frequencies of their speech counterparts, these synthetic sounds follow very closely the spectral and timing changes of speech. The average pitch of the synthetic sounds was 227 Hz, and the average duration was 589 ms. Stimulus pitch did not differ between sound types, $t(8) = 2.2$, *ns*, but synthetic sounds were significantly longer than rhesus monkey calls, $t(8) = 4.1$, $p < .01$ (for a complete description of the stimuli, please see Vouloumanos & Werker, 2004). Consistent with previous studies with similar synthetic sounds (Remez & Rubin, 1993), adults in pilot studies did not report perceiving these synthetic sounds as speech.

Apparatus and Procedure

These were identical to Experiment 1. The presentation of rhesus calls and synthetic tokens was alternated every minute, with order counterbalanced across participants, and the 8 min postbaseline submitted to analysis. Within each minute, tokens were presented randomly with replacement.

Results and Discussion

Neonates preferentially listened to rhesus monkey calls, maintaining their HA sucking for rhesus calls, $t(15) = 2.01$, *ns*, while decreasing their HA sucking for complex synthetic sounds, $t(15) = 2.64$, $p = .038$, over the two observation blocks. Eleven of the 16 neonates exhibited a preference for rhesus monkey calls (see Figure 2). As a final direct test of listening preferences, we compared the relative change in HA sucking across the three studies: Speech versus rhesus calls (Experiment 1), speech versus complex synthetic sounds (Vouloumanos & Werker, 2007), and rhesus calls versus complex synthetic sounds (Experiment 2). In planned comparisons, we found that the HA sucking pattern of neonates' privileging rhesus calls over complex synthetic calls was similar to that of neonates privileging speech over complex synthetic calls (one-tailed, *ns*), and different from neonates that listened to speech and monkey calls equivalently (one-tailed, $p < .05$; see Table 1).

Neonates responded preferentially to speech and to rhesus calls over complex synthetic sounds but showed no preference for speech over rhesus calls. Together, these studies suggest that neonates privilege both speech and calls of rhesus monkeys. The auditory biases of human neonates appear more broadly tuned to properties that are shared by the vocalizations of nonhuman primates.

Experiment 3

To examine whether infants' listening preferences change during early development, we tested 3-month-olds on their preference for speech over rhesus calls. Three-month-old infants were presented with nonsense speech sounds and rhesus monkey calls in alternating trials using an infant-controlled sequential looking preference procedure (Cooper & Aslin, 1990; Vouloumanos & Werker, 2004) in which infants could use their looking behavior to control the presentation of sounds. We measured infants' average looking times for the two types of vocalizations.

Method

Participants

Sixteen 3-month-old infants ($M = 100$ days, age range = 87–111 days) completed the study. An additional 20 infants were excluded as a result of fussing or crying (6), inattention to the screen (5), difficulty seeing infants' eyes (2), failure to look away from the screen to end the trial before the maximum trial length of 60 s had been reached for three or more of the test trials (2), or experimenter error (5). Parents gave informed consent on behalf of their infants. All procedures had been approved by the university's Research Ethics Board.

Stimuli

The auditory stimuli were the same 10 human nonsense speech tokens and 10 rhesus monkey calls that were used in Experiment 1.

Materials and Procedure

Infants were tested using an infant-controlled sequential looking preference procedure (Cooper & Aslin, 1990; Vouloumanos & Werker, 2004) in which infants controlled trial onset and offset by looking at or away from a monitor displaying a fixation checkerboard. Testing was conducted in a sound-attenuated room where infants were seated on a parent's lap in front of a computer monitor. Parents wore headphones playing music to mask the presentation of the experimental sounds. Sounds were played with an average amplitude of 60 dB (±5 dB) from speakers placed below the monitor. Following a musical pretest, infants heard speech and rhesus calls in alternating trials, with order counterbalanced across participants, for a total of eight trials, four trials per sound type.

Infants' attention was drawn to the screen by a red flashing light at the beginning of the experiment and between every trial. Once the infant fixated on the screen, a black and white checkerboard was presented in tandem with one of the sets of sounds. Presentation ceased when the infant looked away from the monitor for a consecutive 2-s period or the maximum trial length of 60 s had been reached. Once the infant looked back at the screen, presentation of the checkerboard recommenced with a new set of sounds. Within each trial, tokens were presented in semirandom order, such that no tokens were ever repeated consecutively. All infants heard a Bach violin concerto in a pretest before the experimental trials began to familiarize them with the procedure. Infants' average looking times for human speech and rhesus monkey calls were calculated based on frame-by-frame offline coding (30 frames/s) of infant looks toward the monitor during each test trial by a coder blind to the experimental condition being tested. Reliability coding performed on 25% of the trials was extremely high (Pearson's $R = .999$, $p < .01$). Three infants were missing values for one trial within a set of four human or four monkey vocalizations; these values were replaced by taking an average of the remaining 15 infants' looking times for that particular sound trial.

Results and Discussion

As there were no significant main or interactive effects of infant sex or order on looking time ($F < 1$, *ns*), we collapsed across these between-subject factors and performed a repeated measures ANOVA on infant looking time with sound type (speech, monkey calls) and trials (1, 2, 3, 4) as within-subject factors. Results revealed a significant main effect of sound type, with 3-month-olds looking longer for speech ($M = 14.6$ s, $SE = 2.01$) compared with rhesus calls ($M = 11.5$ s, $SE = 1.67$), $F(1, 15) = 8.49$, $p = .01$, as well as a main effect of trial, $F(3, 45) = 5.13$, $p < .01$, with overall looking times decreasing in later trials. Twelve of the sixteen 3-month-olds looked longer for speech, as compared to 13 of 30 neonates, Pearson's chi-square $\chi^2(1, N = 46) = 4.22$, $p < .05$ (see Figure 3), confirming a developmental difference in infants' preference for speech relative to rhesus monkey calls.

By at least the age of 3 months, human infants prefer listening to speech compared to rhesus calls. Rapidly in early human development, the relatively broad tuning of neonatal listening preferences is sharpened, becoming selective for speech.

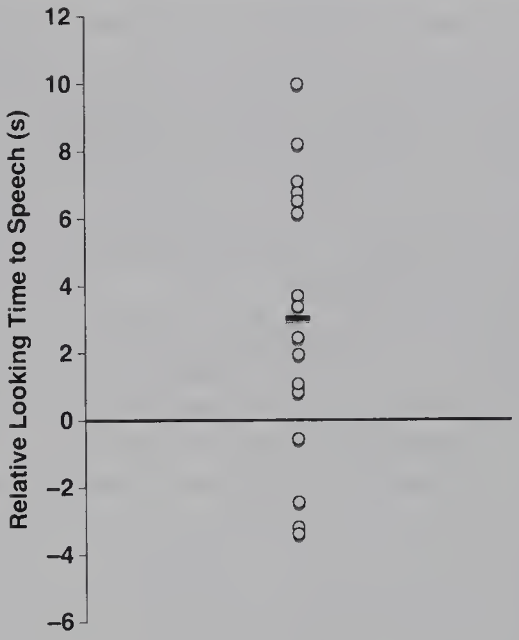


Figure 3. Scatterplot of individual 3-month-olds' relative looking time for speech compared to rhesus monkey vocalizations.
Note. We plot as positive the relative change for speech over monkey calls.

General Discussion

In contrast to previous studies suggesting that humans are born with a listening preference for speech (Vouloumanos & Werker, 2007), these results demonstrate that neonates' listening biases privilege human speech and the calls of another primate, the rhesus macaque. Infants' initial, relatively broad, tuning sharpens rapidly over the course of the first 3 months, resulting in a preference for speech over rhesus calls.

The lack of preference for speech over rhesus monkey calls in neonates is striking, especially when considered in light of previous studies on auditory preferences in human neonates. Neonates typically show preferences for specific speech sounds heard prenatally: They prefer listening to their mother's voice compared to the voice of other females (DeCasper & Fifer, 1980), privilege the rhythmic pattern of their native language over that of unfamiliar languages (Mehler et al., 1988; Moon et al., 1993), and show postnatal recognition of specific songs and stories that had been read to them prenatally (DeCasper & Spence, 1986). These prior findings indicate that the human womb is permeable to sound and, crucially, that the auditory system of the fetus can perceive, encode, and remember some of the specific features of the sounds that are available in the prenatal environment. Given infants' prenatal exposure to speech and their capacity to hear in the third trimester, as well as our selection of speech sounds with attractive intonation patterns (Cooper, Abraham, Berman, & Staska, 1997; Fernald, 1985), one might have expected neonates to prefer speech to monkey calls. Instead, the comparison between speech and monkey calls—as well as the experimental condition confirming a preference for rhesus monkey calls over complex synthetic sounds—demonstrates that speech and rhesus monkey calls can be similarly engaging to newborns.

Why do neonates fail to show a preference for relatively more familiar speech sounds when these are compared to rhesus monkey calls given that they recognize and prefer familiar native and maternal speech vocalizations when various *speech* sounds are contrasted? Previous speech preference studies with newborns used full sentences composed of connected speech (DeCasper & Fifer, 1980; Mehler et al., 1988; Moon et al., 1993). In contrast, our preference studies used isolated words (as in Vouloumanos & Werker, 2004, 2007). Compared to single nonsense words, the connected speech of sentences is longer and, partly as a function of greater length, conveys more of the acoustic (e.g., syllable length, formant structure) and linguistic (e.g., rhythmic, intonational) properties of speech. Sentences may thus allow neonates to extract more information about the signal, including acoustic cues to maternal voice and native language. A comparison of connected speech and rhesus monkey vocalizations could thus reveal a neonatal preference for speech, but isolated words do not.

Neonates may have a general interest in the acoustic properties of voices, harmonically rich sounds composed of quasiperiodic waves punctuated by spectral and temporal irregularities (e.g., Belin, 2006). If this were the case, any vocalization produced by a range of species may engage newborns. More narrowly, neonatal listening preferences may extend to acoustic properties that are shared by other nonhuman primates or may be restricted to vocalizations that show more humanlike spectral filtering. The vocal repertoire of rhesus monkeys is certainly not unique among nonhuman primates, as many other monkeys and apes demonstrate both spectral and temporal variation, with resulting differences in affective and semantic information (Arnold & Zuberbuhler, 2006; Owren, Seyfarth, & Cheney, 1997; Rendall, Seyfarth, Cheney, & Owren, 1999; Zuberbuhler, Cheney, & Seyfarth, 1999). More importantly for this study, rhesus monkeys and humans show similarities in the production mechanisms of their vocalizations. Like humans, rhesus monkeys may be able to control the laryngeal source of their calls independently of the supralaryngeal filter, suggesting possible similarities between human and rhesus neuromotor organization (Hauser & Schon-Ybarra, 1994; Hauser et al., 1993; Rendall et al., 1998). Although few studies have directly examined how humans perceive rhesus monkey calls (cf. Belin et al., 2008; Lewkowicz & Ghazanfar, 2006; Owren & Rendall, 2003), the superficial similarities between the two species' vocalizations may have been sufficient to engage neonates.

What accounts for the developmental shift in infant listening preferences between birth and 3 months? One possibility is that newborns simply could not discriminate between speech and rhesus monkey calls, whereas 3-month-olds did discriminate and therefore preferred listening to speech. Although this possibility cannot be ruled out, it seems unlikely because newborns robustly discriminate much more subtle acoustic differences, for example, between [ba] and [da] (Bertoncini et al., 1987) and between [taep] and [paet] (Moon, Bever, & Fifer, 1992). Our results are thus most likely explained by neonates' engagement with the sounds rather than a failure to discriminate.

A second possibility is that 3-month-olds' preference for speech is a product of increased experience with specific properties of English. Although the increased experience is undeniable, other studies have shown that infants do not recognize native language sound sequences or stress until 9 months of age (Jusczyk, Cutler, & Redanz, 1993; Jusczyk, Friederici, Wessels, & Svenkerud, 1993) and, crucially, do not become attuned to the specific speech sounds of their native language until 6–10 months (Kuhl et al., 1992; Werker & Tees, 1984). Given this relatively late tuning (6–10 months) to native language sound structure, it seems unlikely that the speech preference observed in these relatively younger 3-month-olds in our studies derived specifically from properties of English. Instead, accrued exposure to speech of any kind, or human vocalizations in general, is more likely to contribute to the sharpening of infant preferences.

Finally, the developmental shift may also reflect a shift in infants' attention to different properties of sounds. By 3 months, infants may start to attend to some characteristic properties of speech, such as the alternations between rapid frequency transitions and steady-state frequencies that characterize consonant–vowel alternations (Ladefoged, 2006). Concurrently or alternatively, infants may attend to the specific frequency range (50–5000 Hz) that is characteristic of the human voice. In this latter case, neutral and positively valenced vocalizations that are recognized by infants as human productions, including yawns and laughter, would trigger infants' attention.

The observed shift from more general to species-specific processing in the vocal domain parallels recent results on infant face processing and cross-modal matching: Infants begin life without a human-specific face bias but by 9 months show better processing of human faces compared to the faces of other species (Lewkowicz & Ghazanfar, 2006; Pascalis et al., 2002). Such tuning mechanisms have been observed not only for faces but for language as well. Infants' initial language-general discrimination abilities become language specific by the end of their 1st year (Kuhl et al., 1992; Werker & Tees, 1984). It seems likely that tuning mechanisms are at work in early infancy, sharpening broad initial biases into more specific ones in many domains, and that this may represent a general principle of early perceptual development (Scott, Pascalis, & Nelson, 2007).

Auditory tuning to conspecifics, observed here by 3 months, appears earlier than visual tuning to conspecifics, apparent only at 9 months (Lewkowicz & Ghazanfar, 2006; Pascalis et al., 2002). This is likely because of a combination of maturational and experiential differences in the two systems. Specifically, whereas the auditory system appears functional and responsive by the 6th month of gestation (Gerhardt & Abrams, 2000; Kisilevsky et al., 2003), the visual system, although functional before birth, can only respond to external visual stimuli such as faces after birth. Humans become tuned to conspecific auditory stimuli, and speech in particular, earlier and on a faster time scale (6 vs. 9 months of relevant experience) than they become tuned to conspecific faces. This may very well be because speech, the communicative signal of human conspecifics, *is* special, and rapid tuning reflects this special status.

References

Altmann, C. F., Doehrmann, O., & Kaiser, J. (2007). Selectivity for animal vocalizations in the human auditory cortex. *Cerebral Cortex*, *17*, 2601–2608.

Arnold, K., & Zuberbuhler, K. (2006). Language evolution: Semantic combinations in primate calls. *Nature*, *441*(7091), 303.

Belin, P. (2006). Voice processing in human and nonhuman primates. *Philosophical transactions of the Royal Society of London Series B: Biological Sciences*, *361*(1476), 2091–2107.

Belin, P., Fecteau, S., Charest, I., Nicastro, N., Hauser, M. D., & Armony, J. L. (2008). Human cerebral response to animal affective vocalizations. *Proceedings of the Royal Society of London Series B, Biological Sciences*, *275*(1634), 473–481.

Bertoncini, J., Bijeljac-Babic, R., Blumstein, S. E., & Mehler, J. (1987). Discrimination in neonates of very short CVs. *Journal of the Acoustical Society of America*, *82*, 31–37.

Bertoncini, J., Bijeljac-Babic, R., Jusczyk, P. W., Kennedy, L. J., & Mehler, J. (1988). An investigation of young infants' perceptual representations of speech sounds. *Journal of Experimental Psychology: General*, *117*, 21–33.

Butterfield, E. C., & Siperstein, G. N. (1970). Influence of contingent auditory stimulation upon non-nutritional suckle. In J. F. Bosma (Ed.), *Third symposium on oral sensation and perception: The mouth of the infant* (pp. 313–334), Springfield, IL: Charles C. Thomas.

Cooper, R. P., Abraham, J., Berman, S., & Staska, M. (1997). The development of infants' preference for motherese. *Infant Behavior & Development*, *20*, 477–488.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, *61*, 1584–1595.

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, *208*(4448), 1174–1176.

DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior & Development*, *9*, 133–150.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*(968), 303–306.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior & Development*, *8*, 181–195.

Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *Journal of the Acoustical Society of America*, *102*, 1213–1222.

Fitch, W. T. (2000). The evolution of speech: A comparative review. *Trends in Cognitive Sciences*, *4*, 258–267.

Fitch, W., & Hauser, M. D. (1995). Vocal production in nonhuman primates: Acoustics, physiology, and functional constraints on "honest" advertisement. *American Journal of Primatology*, *37*, 191–219.

Floccia, C., Nazzi, T., & Bertoncini, J. (2000). Unfamiliar voice discrimination for short stimuli in newborns. *Developmental Science*, *3*, 333–343.

Gerhardt, K. J., & Abrams, R. M. (2000). Fetal exposures to sound and vibroacoustic stimulation. *Journal of Perinatology*, *20*(8, Pt. 2), S21–S30.

Gottlieb, G. (1997). *Synthesizing nature-nurture: Prenatal roots of instinctive behavior*. Mahwah, NJ: Erlbaum.

Hauser, M. D. (1992). Articulatory and social factors influence the acoustic structure of rhesus monkey vocalizations: A learned mode of production? *Journal of the Acoustical Society of America*, *91*, 2175–2179.

Hauser, M. D. (1996). Vocal communication in *Macaques*. In J. E. Fa & D. G. Lindburg (Eds.), *Evolution and ecology of macaque societies* (pp. xvii). Cambridge, UK: Cambridge University Press.

Hauser, M. D. (2000). A primate dictionary? Decoding the function and meaning of another species' vocalizations. *Cognitive Science*, *24*, 445–475.

Hauser, M. D., Evans, C. S., & Marler, P. (1993). The role of articulation in the production of rhesus monkey, *Macaca mulatta*, vocalizations. *Animal Behaviour*, *45*, 423–433.

Hauser, M. D., & Fowler, C. A. (1992). Fundamental frequency declination is not unique to human speech: Evidence from nonhuman primates. *Journal of the Acoustical Society of America*, *91*, 363–369.

Hauser, M. D., & Marler, P. (1993). Food-associated calls in rhesus macaques (*Macaca mulatta*): I. Socioecological factors. *Behavioral Ecology*, *4*, 194–205.

Hauser, M. D., & Schon-Ybarra, M. (1994). The role of lip configuration in monkey vocalizations: Experiments using xylocaine as a nerve block. *Brain and Language*, *46*, 232–244.

Johnson, M. H., Bolhuis, J. J., & Horn, G. (1985). Interaction between acquired preferences and developing predispositions during imprinting. *Animal Behaviour*, *33*, 1000–1006.

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, *64*, 675–687.

Jusczyk, P. W., Friederici, A. D., Wessels, J. M., & Svenkerud, V. Y. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, *32*, 402–420.

Kisilevsky, B. S., Hains, S. M., Lee, K., Xie, X., Huang, H., Ye, H. H., et al. (2003). Effects of experience on fetal voice recognition. *Psychological Science*, *14*, 220–224.

Kuhl, P. K., Coffey-Corina, S., Padden, D., & Dawson, G. (2005). Links between social and linguistic processing of speech in preschool children with autism: Behavioral and electrophysiological measures. *Developmental Science*, *8*, F1–F12.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606–608.

Ladefoged, P. (2006). *A course in phonetics* (5th ed.). Boston: Thomson Wadsworth.

Lewkowicz, D. J., & Ghazanfar, A. A. (2006). The decline of cross-species intersensory perception in human infants. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 6771–6774.

Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, *4*, 187–196.

Marler, P., & Peters, S. (1977). Selective vocal learning in a sparrow. *Science*, *198*(4316), 519–521.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*, 143–178.

Moon, C., Bever, T. G., & Fifer, W. P. (1992). Canonical and non-canonical syllable discrimination by two-day-old infants. *Journal of Child Language*, *19*, 1–17.

Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior & Development*, *16*, 495–500.

Nazzi, T., Floccia, C., & Bertoncini, J. (1998). Discrimination of pitch contours by neonates. *Infant Behavior & Development*, *21*, 779–784.

Nicastro, N., & Owren, M. J. (2003). Classification of domestic cat (*Felis catus*) vocalizations by naive and experienced human listeners. *Journal of Comparative Psychology*, *117*, 44–52.

Owren, M. J., & Rendall, D. (2003). Salience of caller identity in rhesus monkey (*Macaca mulatta*) coos and screams: Perceptual experiments with human (*Homo sapiens*) listeners. *Journal of Comparative Psychology*, *117*, 380–390.

Owren, M. J., Seyfarth, R. M., & Cheney, D. L. (1997). The acoustic features of vowel-like grunt calls in chacma baboons (*Papio cyncephalus ursinus*): Implications for production processes and functions. *Journal of the Acoustical Society of America*, *101*, 2951–2963.

Pascalis, O., de Haan, M., & Nelson, C. A. (2002). Is face processing species-specific during the first year of life? *Science*, *296*(5571), 1321–1323.

Peña, M., Maki, A., Kovacic, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, F., et al. (2003). Sounds and silence: An optical topography study of language recognition at birth. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 11702–11705.

Pinker, S., & Jackendoff, R. (2005). The faculty of language: What's special about it? *Cognition*, *95*, 201–236.

Remez, R. E., & Rubin, P. E. (1993). On the intonation of sinusoidal sentences: Contour and pitch height. *Journal of the Acoustical Society of America*, *94*, 1983–1988.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science*, *212*(4497), 947–949.

Rendall, D., Owren, M. J., & Rodman, P. S. (1998). The role of vocal tract filtering in identity cueing in rhesus monkey (*Macaca mulatta*) vocalizations. *Journal of the Acoustical Society of America*, *103*, 602–614.

Rendall, D., Seyfarth, R. M., Cheney, D. L., & Owren, M. J. (1999). The meaning and function of grunt variants in baboons. *Animal Behaviour*, *57*, 583–592.

Scott, L. S., Pascalis, O., & Nelson, C. A. (2007). A domain-general theory of the development of perceptual discrimination. *Current Directions in Psychological Science*, *16*, 197–201.

Smith, E. C., & Lewicki, M. S. (2006). Efficient auditory coding. *Nature*, *439*(7079), 978–982.

Spence, M. J., & DeCasper, A. J. (1987). Prenatal experience with low-frequency maternal-voice sounds influence neonatal perception of maternal voice samples. *Infant Behavior & Development*, *10*, 133–142.

Vouloumanos, A., & Werker, J. F. (2004). Tuned to the signal: The privileged status of speech for young infants. *Developmental Science*, *7*, 270–276.

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science*, *10*, 159–164.

Walker-Andrews, A. S. (1997). Infants' perception of expressive behaviors: Differentiation of multimodal information. *Psychological Bulletin*, *121*, 437–456.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development*, *7*, 49–63.

Zuberbuhler, K., Cheney, D. L., & Seyfarth, R. M. (1999). Conceptual semantics in a nonhuman primate. *Journal of Comparative Psychology*, *113*, 33–42.

Child Development, March/April 2010, Volume 81, Number 2, Pages 528–539

Early Gesture Predicts Language Delay in Children With Pre- or Perinatal Brain Lesions

Eve Sauer, Susan C. Levine, and Susan Goldin-Meadow
University of Chicago

Does early gesture use predict later productive and receptive vocabulary in children with pre- or perinatal unilateral brain lesions (PL)? Eleven Children with PL were categorized into 2 groups based on whether their gesture at 18 months was within or below the range of typically developing (TD) children. Children with PL whose gesture was within the TD range developed a productive vocabulary at 22 and 26 months and a receptive vocabulary at 30 months that were all within the TD range. In contrast, children with PL below the TD range did not. Gesture was thus an early marker of which children with early unilateral lesions would eventually experience language delay, suggesting that gesture is a promising diagnostic tool for persistent delay.

Children with pre- or perinatal unilateral brain lesions (PL) exhibit marked plasticity for language functions. Even when their lesions affect classic language areas, children typically do not exhibit the aphasias that adults with similar lesions display (e.g., Bates & Dick, 2002; Feldman, 2005; Levine, Kraus, Alexander, Suriyakham, & Huttenlocher, 2005; Reilly, Levine, Nass, & Stiles, 2008; Stiles, Reilly, Paul, & Moses, 2005; Woods & Teuber, 1978). However, children with PL often exhibit delays in both productive and receptive language (e.g., Bates, Thal, & Janowsky, 1992; Chilosi, Cipriani, Bertuccelli, Pfanner, & Cioni, 2001; Feldman, Holland, Kemp, & Janosky, 1992; Marchman, Miller, & Bates, 1991; Thal et al., 1991). Importantly, these delays are transient for some children with PL but persistent for others (Bates et al., 1997; Feldman et al., 1992; Thal et al., 1991; Vicari et al., 2000). The question we ask here is whether early child gesture can be used to predict subsequent vocabulary development in children with PL, just as it can for typically developing (TD) children (Rowe & Goldin-Meadow, 2009a). If so, gesture may be a potential early diagnostic indicator of which children with PL will exhibit subsequent language delays and which will catch up and thus fall within the normative range.

The relation between early gesture and vocabulary growth has been extensively studied in TD children. Many children begin by communicating through gesture and over time replace gesture with speech (Bates, Benigni, Bretherton, Camaioni, & Volterra, 1979; Bates & Dick, 2002; Iverson, Capirci, & Caselli, 1994). Early gesture use not only precedes subsequent vocabulary development, it also predicts subsequent vocabulary production and comprehension (Goldin-Meadow & Butcher, 2003; Iverson & Goldin-Meadow, 2005; O'Reilly, Painter, & Bornstein, 1997; Rowe & Goldin-Meadow, 2009a, 2009b; Rowe, Özçaliskan, & Goldin-Meadow, 2006, 2008). For example, Iverson and Goldin-Meadow (2005) were able to predict which lexical items entered a child's spoken vocabulary from looking at the child's gestures several months earlier. Specifically, they found that children produced a deictic pointing gesture to refer to an object, person, or location (e.g., point at a ball) approximately 3 months before producing a word for that object, person, or location (e.g., the word *ball*). As a second example, Rowe et al. (2008) found that the number of different meanings children conveyed with their gestures at 14 months was significantly related to the size of their receptive vocabularies at 42 months (see also Rowe & Goldin-Meadow, 2009a, 2009b).

This research was supported by P01HD40605 from NICHD and a grant from the Brain Research Foundation. We thank the participating families for sharing their child's language development with us; K. Brasky, L. Chang, E. Croft, K. Duboc, J. Griffin, S. Gripshover, K. Harden, L. King, C. Meanwell, E. Mellum, M. Nikolas, J. Oberholtzer, C. Rousch, L. Rissman, B. Seibel, M. Simone, K. Uttich, and J. Wallman for help in collecting and transcribing the data; Seyda Özçalıskan for gesture coding; J. Fisher, P. Huttenlocher, S. Small, and M. Staudt for coding brain scans; Meredith Rowe for statistical advice; and Kristi Schonwald and Jason Voigt for administrative and technical assistance.

Correspondence concerning this article should be addressed to Eve Sauer, Department of Psychology, University of Chicago, 5848 South University Avenue, Chicago, IL 60637. Electronic mail may be sent to easauer@uchicago.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8102-0010

This relation held even after controlling for the number of different spoken words that the children produced at 14 months.

Previous research has examined gesture use in children with PL but has not directly related it to language delay. For example, Bates et al. (1997) found that children with early right hemisphere lesions show delays in gesture between 10 and 17 months of age, based on parent responses on the MacArthur Communicative Development Inventory (MCDI; see also Marchman et al., 1991). Are the children who exhibit delays in gesture the same children who exhibit delays in vocabulary development? We might expect delays in gesture use to go hand in hand with language delays simply because gesture and language form an integrated system, not only in adults (McNeill, 1992) but also in TD children at the early stages of language learning (Goldin-Meadow, 2003). If the gesture–language system is robust in the face of early unilateral brain injury, children whose language development is proceeding at a typical pace should display typical gesture, and children whose language is delayed should display delays in gesture. Moreover, early gesture should predict subsequent language development as it does in TD children.

It is, of course, possible that gesture and language do *not* form an integrated system in children with PL. Early brain injury could disrupt the relation between language (both receptive and expressive) and gesture, leading to different patterns of gesture–language development than those observed in TD children. In addition, children with PL frequently have motor impairments involving the contralesional hand and arm (Huttenlocher, 2002; Levine, Huttenlocher, Banich, & Duda, 1987; Mercuri et al., 2004; Staudt et al., 2002). Such impairments might limit the use of gesture in early language development and thus disrupt the gesture–language link in children with PL.

The goal of our study is to explore whether early child gesture predicts later vocabulary size in children with PL as it does in TD children. If so, not only will we have evidence that gesture and language form a tightly organized system that can withstand early injury to the brain, but we also will have uncovered an early marker that might be used to identify those children with PL who are at risk for subsequent language delay and who might benefit from early intervention.

Method

Participants

Participants were 11 children with PL (8 girls, 3 boys) studied longitudinally between 18 and 30 months. For 7 of the 11 children, we also have data about language comprehension based on parental responses on the MCDI (Fenson et al., 1993). Lesion characteristics of the individual children are given in Table 1; for 10 of the 11 children, these characteristics were based on magnetic resonance imaging (MRI), and for the remaining child, the characteristics were based on detailed medical notes from previous CT and MRI scans. Two pediatric neurologists reviewed all of the information available and agreed that each of the 11 children has a unilateral lesion. The lesions were either the result of cerebrovascular infarcts of the middle cerebral artery territory (CV lesions), primarily

Table 1
Lesion Characteristics for Individual Children With Pre- or Perinatal Unilateral Brain Lesions

ID	Sex	Hemisphere	Size	Type	Location
1	Female	Left	Large	Cerebrovascular infarct	F, T, P, O, subcortical
2	Female	Right	Large	Cerebrovascular infarct	F, T, P, subcortical
3	Male	Right	Small	Periventricular	Subcortical
4	Female	Left	Large	Cerebrovascular infarct	F, T, P, subcortical
5	Male	Left	Medium	Cerebrovascular infarct	F, T, P, subcortical
6	Male	Left	Small	Periventricular	F, T, subcortical
7	Female	Left	Large	Cerebrovascular infarct	F, T, P, O subcortical
8	Female	Left	Large	Cerebrovascular infarct	F, T, P
9	Female	Right	Large	Cerebrovascular infarct	F, T, P, O, subcortical
10	Female	Left	Small	Periventricular	T, subcortical
11	Female	Left	Small	Periventricular	Details not available

Note. F = frontal, T = temporal, P = parietal, O = occipital, subcortical = involves subcortical areas.

affecting inferior frontal and superior temporal regions, or the result of periventricular lesions (PV lesions), primarily affecting subcortical structures and white matter tracts. Periventricular lesions in very low birthweight, prematurely born children have been the focus of much previous literature (although see Krägeloh-Mann & Horber, 2007, for a review of recent studies of PV lesion in full-term infants). Our sample is more restricted because we excluded children born earlier than 36 weeks gestation. Children were recruited through area health providers and parent support groups, and came from middle- to upper-middle-class Caucasian families. All children were being raised as monolingual English speakers.

Lesions were characterized by type (CV or PV), laterality (left or right), and size (small, medium, or large). Lesion size was determined using the following criteria: Small lesions affected only one lobe or minimally affected subcortical regions, medium lesions affected more than one lobe or extended into more than one subcortical region, and large lesions affected three or four lobes and often affected subcortical regions and the thalamus (large lesions were typically cerebrovascular infarcts). Four children with PL had small lesions, 1 had a medium lesion, and 6 had large lesions; we collapsed children with small and medium lesions into a single group based on previous findings indicating that the language development of these groups tend to pattern together (Brasky, Nikolas, Meanwell, Levine, & Goldin-Meadow, 2005). Four of the children had periventricular lesions (PV) and 7 had cerebrovascular infarcts (CV). Eight children had left hemisphere lesions and 3 had right hemisphere lesions.

To situate the gesture and speech development of children with PL within a normative sample, we also observed 53 TD children (26 girls, 27 boys) between 18 and 26 months whose gesture and speech development have been previously described (Rowe et al., 2006, 2008; Rowe & Goldin-Meadow, 2009a, 2009b). All TD children were being raised as monolingual English speakers and had no known medical conditions. The families of the TD children reflected the demographic diversity of the Chicago area in terms of family income and ethnicity, and thus contained families of low, middle, and high socioeconomic status (SES). In contrast, the families of the children with PL were skewed toward the high end of the SES continuum. Because children from higher SES families tend to have larger vocabularies than children from lower SES families (Hart & Risley, 1995), the SES difference between the two groups would work against our finding language delays in the children with PL. Furthermore, findings from Rowe et al. (2008) suggest that early gesture predicts subsequent vocabulary comprehension in TD children regardless of SES. Thus, the SES difference between the two groups is not likely to affect the pattern of results.

Procedures

All children were observed in their homes for 90 min while interacting naturally with their primary caregivers at 18, 22, and 26 months. Parents were asked to perform their everyday activities and interact with their children as they typically would. The Peabody Picture Vocabulary Test–Third Edition (PPVT–3; Dunn & Dunn, 1997), a normed test of vocabulary comprehension, was given to the children at 30 months.

Coding Speech and Gesture

All child speech produced during each videotaped session was transcribed. Sounds that were reliably used to refer to entities, properties, or events (e.g., ''doggie,'' ''nice,'' ''broken''), along with onomatopoeic sounds (e.g., ''meow,'' ''choo-choo'') and conventionalized evaluative sounds (e.g., ''oopsie,'' ''uh-oh''), were counted as words. We established reliability for our speech transcripts by having a second individual transcribe 20% of the videotapes; reliability was achieved when the second coder agreed with the first on 95% of the transcription decisions.

Any communicative hand movement that did not involve direct manipulation of objects (e.g., twisting a jar open) or a ritualized game (e.g., patty cake) was considered a gesture. Each gesture was classified as deictic, iconic, or conventional: (a) *Deictic gestures* were those that indicated concrete objects, people, or locations. We considered these objects, people, and locations to be the referents of the deictic gestures (e.g., pointing to a dog referred to a *dog*, holding up a bottle referred to a *bottle*). (b) *Iconic gestures* were those that depicted the attributes or actions of an object via hand or body movements (e.g., moving the index finger in circles to convey a ball *rolling*). (c) *Conventional* gestures were those whose forms and meanings are prescribed by the culture (e.g., nodding the head to mean *yes*, extending an open palm next to a desired object to mean *give*). We assessed reliability for gesture on a subset of the videotaped sessions coded by an independent coder. Agreement between coders was 88% (κ = .76; $N = 763$) for identifying gestures and 91% (κ = .86; $N = 375$) for assigning meanings to gestures.

Measures

We calculated the number of unique spoken words (speech types) that each child produced during a single session and used this measure as an index of the child's productive vocabulary in speech at each age. Several decisions were made as to what constituted a word type. Morphologically inflected variants of words (e.g., *run*, *running*) were considered a single type. Words produced in imitation of the mother were included in the corpus of child word types, as were words that the child produced in the context of book reading. We included imitations because children typically only imitate linguistic structures that are within their spontaneous language reach (Slobin & Welsh, 1967). Furthermore, the children with PL produced a very small number of imitations ($M = 1.09$, accounting for only 1.53% of their total speech types), as did the TD children ($M = 1.42$, accounting for 1.9% of their total speech types). Thus, it is likely that including imitations has little effect on the results. The number of word types produced by each child at each session served as our measure of productive vocabulary.

The PPVT–3 (Dunn & Dunn, 1997), administered at 30 months, served as our measure of children's receptive vocabulary. We chose this measure because it is widely used, sensitive to individual differences, standardized, and can be used reliably with children as young as 30 months.

We calculated the number of unique gesture meanings (gesture types) that each child produced at each session. We counted gestures that conveyed different meanings as unique gesture types. The meaning assigned to a deictic gesture was the object, person, or place to which the deictic gesture referred. Note that under this system, each pointing gesture indicating a different entity is considered a unique type. For example, if a child pointed to a ball and a cup, the child would be given credit for two gesture types in his repertoire, *ball* and *cup*. This procedure for attributing meaning to early child gesture has been used in previous studies and, importantly, results in systematic patterns. For example, when this system is used to code early child communications in TD children, we are able to predict the size of a child's later receptive vocabulary (Rowe & Goldin-Meadow, 2009a, 2009b) and the onset and nature of the child's two-word speech (Goldin-Meadow & Butcher, 2003; Iverson & Goldin-Meadow, 2005; Özçaliskan & Goldin-Meadow, 2005; see Goldin-Meadow & Mylander, 1984; Iverson & Goldin-Meadow, 2005, for further justification of the coding system, and details of how meanings were assigned to gestures).

Results

Gesture Use From 18 to 26 Months in Children With PL

We first investigated the spontaneous use of gesture in children with PL at 18, 22, and 26 months. When looking at overall amount of gesturing, we find that most of the gestures that the children with PL produced at all three ages were deictic gestures (average across ages = 72%), as were the gestures produced by the TD children (average across ages = 75%) and there was not a significant difference between children with PL and TD children on this measure ($F = 1.05$, $p = .31$). Also, children with PL did not differ significantly from TD children in total number of gestures produced at 18 or 22 months of age (18 months: $t = 1.39$, $p = .17$; 22 months: $t = 0.32$, $p = .75$). But there was a marginally significant difference between the groups at 26 months ($t = 1.86$, $p = .07$), with children with PL producing an average of 75 gestures ($SD = 38.35$) and TD children producing an average of 113 gestures ($SD = 63.96$) during the 90-min session.

We next turn to our measure of gesture use: gesture types. We chose to use gesture types rather than the total number of gestures because the gesture type measure captures the diversity of meanings each child conveyed in gesture and, in this sense, is a close analog of our speech type measure. Moreover, the number of gesture types was highly correlated with the total number of gestures in our sample of children with PL: $\rho = .96$, $p < .001$ at 18 months; $\rho = .93$, $p < .001$ at 22 months; and $\rho = .96$, $p < .001$ at 26 months (Spearman rank-order correlations). Gesture types were predominantly deictic gestures for both children with PL and TD children (PL: 86% average across ages; TD: 82% average across ages). TD children and children with PL did not significantly differ in the proportion of gesture types produced with a deictic gesture ($F = 1.18$, $p = .28$).

As a group, the children with PL produced approximately the same number of gesture types as the TD children. On average, the children with PL produced 28.73 gesture types at 18 months

(*SD* = 21.89), 45.55 gesture types at 22 months (*SD* = 22.94), and 38.64 gesture types at 26 months (*SD* = 18.77). Figure 1 displays the mean number of gesture types that the children with PL produced in relation to the gesture type data from the TD children at each age. The boxes in the graph represent the interquartile range for the 53 TD children; the line in the middle of each box represents the median for the TD children and the tails represent the 5th and 95th percentiles. Note that, as a group, the children with PL were close to the median for the TD children's production of gesture types and *t* tests confirmed that there were no significant differences between groups (18 months: $t = 0.68$, $p = .50$; 22 months: $t = -0.67$, $p = .50$; 26 months: $t = 0.66$, $p = .51$).

There was, however, a great deal of variability within the group of children with PL. To explore this variability, we divided the children with PL into two groups based on their production of gesture types at 18 months: (a) Children in the LOW group fell below the 25th percentile for gesture production at 18 months in the TD group (25th percentile = 14.5 gesture types). Children in the HIGH group fell above the 25th percentile. Thus, we used the 25th percentile as a cutoff to identify children at the low end of the typical range (Thal et al., 1991). Based on these criteria, 5 of the 11 children with PL were classified into the LOW gesture group, and 6 were classified into the HIGH gesture group.

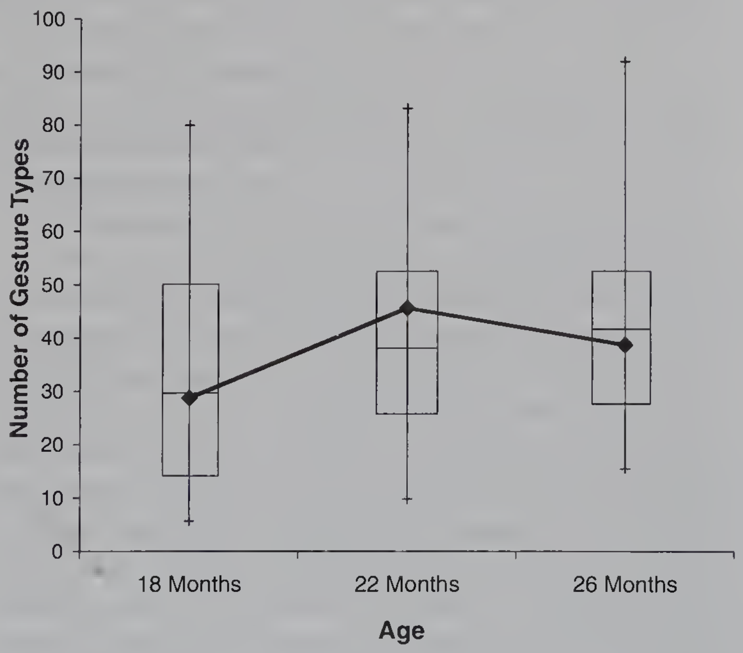


Figure 1. Mean number of gesture types that the children with pre- or perinatal unilateral brain lesions produced at each observation session.
Note. The boxes in the graph represent the interquartile range for gesture types in the typically developing (TD) children; the line in the middle of each box represents the median for TD children, and the tails represent the 5th and 95th percentiles.

Importantly, there was stability in the children's gesture use between 18 and 26 months: All 5 of the children who were assigned to the LOW gesture group on the basis of their 18-month performance fell below the 25th percentile cutoff on at least one of the other two sessions (22 and 26 months); 2 of these children fell below the cutoff on both sessions. Moreover, all 6 of the children assigned to the HIGH gesture group on the basis of their 18-month performance fell above the cutoff on both of the other two sessions. We ran permutation tests at each age (Hesterberg, Moore, Monaghan, Clipson, & Epstein, 2006) to determine whether the number of gesture types produced by the HIGH gesture group differed significantly from the number produced by the LOW gesture group. The HIGH gesture group not only produced significantly more gesture types than the LOW gesture group at 18 months (the age at which children were divided into groups, mean difference = 33.6, $p = .006$), but they also produced more gesture types at 22 months (mean difference = 39.5, $p < .001$) and at 26 months (mean difference = 30.87, $p = .002$; see Figure 2). Thus, early differences in gesture use in the children with PL were stable over time.

Does Gesture Production at 18 Months Predict Later Productive Vocabulary?

We next investigated whether gesture use at 18 months can be used to predict later speech use. On average, the children with PL produced 15.36 (*SD* = 14.64) speech types at 18 months, 56.91 (*SD* = 36.44) at 22 months, and 123.36 (*SD* = 77.23) at 26 months. Thus, as a group, the children with PL increased the number of different words they produced over time. However, there was once again a great deal of variability within the group. Can this variability be related to gesture use at 18 months?

Figure 3 presents the mean number of speech types children in the LOW and HIGH gesture groups produced at each age, displayed in relation to the speech type data from the TD children. We ran permutation tests at 18 months to determine whether the number of speech types produced by the HIGH gesture group differed significantly from the number produced by the LOW gesture group. At 18 months, the age used to divide the children into LOW and HIGH gesture groups, the HIGH group produced slightly more speech types than

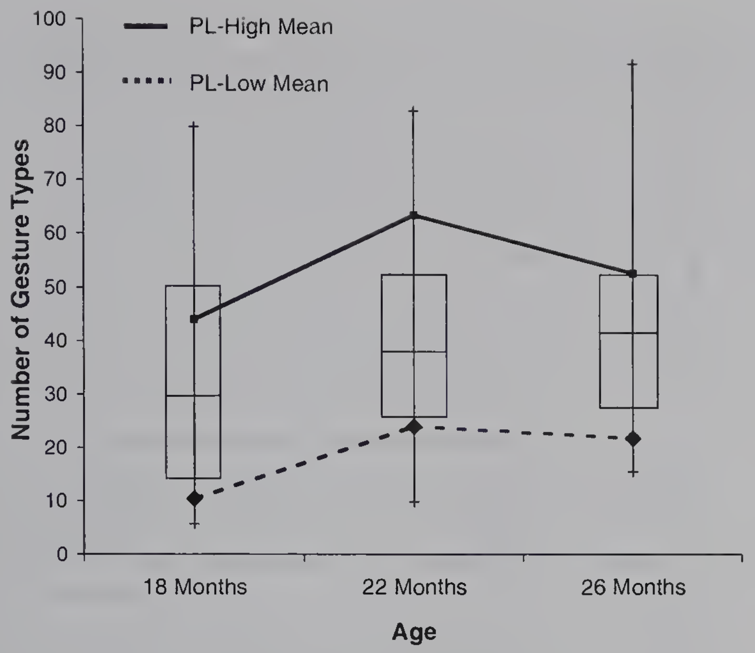


Figure 2. Mean number of gesture types produced by children with pre- or perinatal unilateral brain lesions (PL) in the HIGH and LOW gesture groups at each observation session.
Note. Children with PL were divided into gesture groups based on their production of gesture types at 18 months. The boxes in the graph represent the interquartile range for gesture types in the typically developing (TD) children; the line in the middle of each box represents the median for TD children and the tails represent the 5th and 95th percentiles.

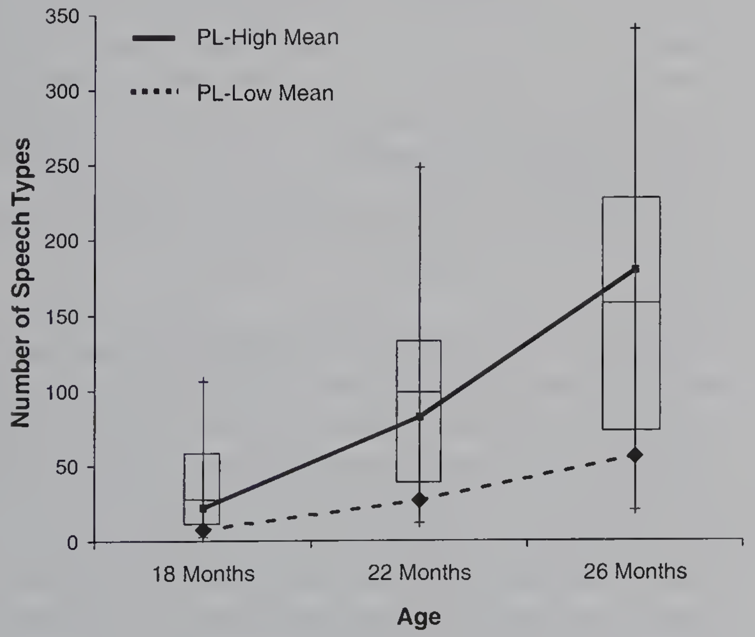


Figure 3. Mean number of speech types produced by children with pre- or perinatal unilateral brain lesions (PL) in the HIGH and LOW gesture groups at each observation session.
Note. Children with PL were divided into gesture groups based on their production of gesture types at 18 months. The boxes in the graph represent the interquartile range for speech types in the typically developing (TD) children; the line in the middle of each box represents the median for the TD children, and the tails represent the 5th and 95th percentiles.

the LOW group, but the difference was not reliable (mean difference = 14.6, $p = .112$). At 18 months, all 5 LOW gesture children fell below the TD 25th percentile for speech types, but 3 of the 6 HIGH gesture children did as well (Fisher exact probability = .12). In other words, at 18 months, it was not possible to reliably distinguish the two groups on the basis of their speech use, only their gesture use.

The interesting question, however, is whether gesture use at 18 months presages *later* speech use. We ran permutation tests at 22 and 26 months to determine whether the number of speech types produced by the HIGH gesture group differed significantly from the number produced by the LOW gesture group and found that they did. The HIGH gesture group reliably produced more speech types than the LOW gesture group at 22 months (mean difference = 55.2, $p = .003$) and at 26 months (mean difference = 123.5, $p = .005$). Moreover, at 22 months, all 5 LOW gesture children continued to fall below the TD 25th percentile for speech types but none of the 6 HIGH gesture children did (Fisher exact probability = .002). At 26 months, 4 of the 5 LOW gesture children fell below the TD 25th percentile for speech types, compared to none of the 6 HIGH gesture children (Fisher exact probability = .01). Although the sample is small, we can use gesture production at 18 months to predict which children will be below the TD 25th percentile in speech production at 22 and 26 months with some precision: At 22 months, positive predictive value for LOW gesture children to be below the TD 25th percentile was 1.00, and sensitivity was 1.00; at 26 months, positive predictive value was .80 and sensitivity was 1.00. We used categorical analyses (rather than regression analyses) to test whether later speech production can be predicted from early gesture use because our data do not meet the assumptions for linear regression. However, it is worth noting that 18-month gesture does significantly correlate with productive vocabulary at both 22 and 26 months (22 months: $\rho = .81$, $p = .002$; 26 months: $\rho = .81$, $p = .003$, Spearman's rank-order correlations).

Does Gesture Production at 18 Months Predict Later Receptive Vocabulary?

We have found thus far that differences in gesture production in children with PL at 18 months predict later differences in *productive* vocabulary. We next ask whether early gesture use also predicts later differences in *receptive* vocabulary, that is, performance on the PPVT–3. In addition to extending our language probes from production to comprehension, using the PPVT–3 has the advantage of allowing us to assess children's performance on a standardized measure of language development. Moreover, the PPVT–3 relies on a different data collection technique from our speech production measure, one that is less tightly tied to our gesture measure. Our speech and gesture production measures both come from spontaneous communication during a 90-min period. In contrast, the PPVT–3 is a normed and validated receptive vocabulary test in which the child is given a word and must select the picture he thinks the word refers to out of a distractor set. The PPVT–3 thus allows us to test the robustness of the relation between gesture and early language.

The children with PL had an average PPVT–3 standard score of 84.64 (*SD* = 21.82). This score is approximately 1 *SD* below the average score for this test (*M* = 100, *SD* = 15; Dunn & Dunn, 1997). Thus, as a group, the children with PL exhibited relatively low vocabulary comprehension. However, there was great variability in the vocabulary comprehension scores of the children with PL, ranging from a low of 45 to a high of 113. The standard deviation for the group of children with PL was 21.8, considerably higher than the standard deviation of 15 in the PPVT–3 norms. Thus, some children with PL were performing well within the average range on the PPVT–3, whereas others were falling far below that range.

Can we predict performance on the PPVT–3 at 30 months from the gestures the children produced a year earlier, at 18 months? On average, children in the LOW gesture group (*M* = 66.8, *SD* = 16.65) scored significantly lower on the PPVT–3 than children in the HIGH gesture group (*M* = 99.5, *SD* = 12.15) (*U* = 1.00, *p* = .009, Mann–Whitney). Moreover, all 5 LOW gesture children fell below the 25th percentile for the PPVT–3, whereas only 2 of the 6 HIGH gesture children did (Fisher exact probability = .045). Thus, positive predictive value for LOW gesture children to be below the 25th percentile on the PPVT–3 was 1.00, and sensitivity was .71.

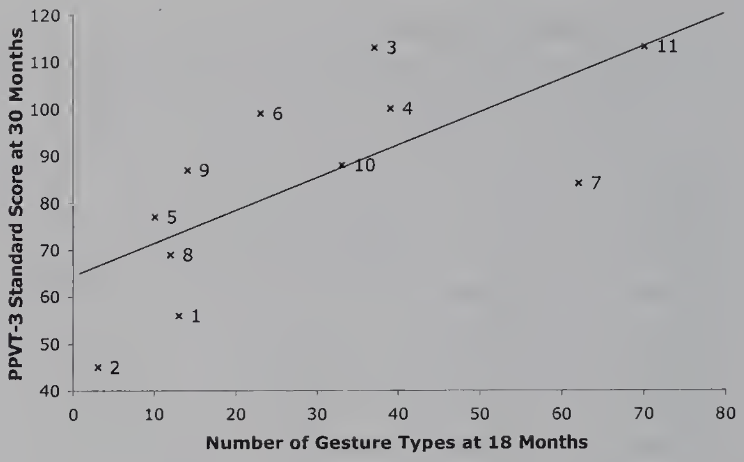


Figure 4. The relation between the number of gesture types produced at 18 months and vocabulary comprehension score on the Peabody Picture Vocabulary Test–Third Edition (PPVT–3) at 30 months in children with pre- or perinatal unilateral brain lesions.

Note. Each x represents an individual. The labels correspond to the ID numbers in Table 1. The five leftmost points (1, 2, 5, 8, 9) represent children in the LOW gesture group. The regression line on the graph is from Model 1.

Table 2

Regression Models Predicting Peabody Picture Vocabulary Test–Third Edition Scores at 30 Months Using Gesture Types and Speech Types at 18 Months in Children With Pre- or Perinatal Unilateral Brain Lesions (n = *11*)

	Vocabulary Comprehension, *B*(*SE*)	
	Model 1	Model 2
Intercept	64.77** (8.50)	61.05** (9.57)
Child gesture types at 18 months	0.69* (0.24)	0.65* (0.25)
Child speech types at 18 months		0.33 (0.37)
R^2 statistic (%)	48.1	52.8

**p* < .05. ***p* < .01.

We also explored the relation between 18-month gesture and 30-month receptive vocabulary using a continuous measure of gesture production. Figure 4 presents a scatter plot of the relation between gesture types at 18 months (*x*-axis) and receptive vocabulary (PPVT–3) at 30 months (*y*-axis); each point represents an individual child. We conducted Spearman rank order correlations between the number of gesture types a child produced at 18 months and that child's PPVT–3 standard score at 30 months, and found a significant relation between the two measures (ρ = .80, *p* = .003). Importantly, the number of speech types a child produced at 18 months did *not* significantly correlate with that child's PPVT–3 standard score at 30 months (ρ = .34, *p* = .31), presumably because there was very little variation in the number of speech types these children produced at 18 months.

We then used a linear regression model to explore whether gesture production at 18 months predicts vocabulary comprehension at 30 months (Table 2). The data meet the assumptions of linear regression required for these analyses: There is a linear relation, the residuals are normally distributed and all are within ±2, and there is no evidence that homeoskedasticity is violated. We found that number of gesture types at 18 months significantly predicted PPVT–3 scores at 30 months, accounting for 48.1% of the variance (Model 1). When we included the number of child speech types at 18 months in the regression model, the number of gesture types at 18 months continued to significantly predict 30-month word comprehension on the PPVT–3 (Model 2); in fact, it was the only significant predictor and Model 2 accounted for 52.8% of the variance in children's PPVT–3 scores. Thus, early child gesture predicts later receptive vocabulary even when controlling for early child speech.

We have found that for children with PL, variability in early gesture use predicts later variability in receptive vocabulary. But is gesture use the only measure that predicts later receptive vocabulary in these children? It is likely that the children also varied in their ability to comprehend spoken language at the early time point. Variability in early language comprehension might then be as good as, or an even better, predictor of later variability in receptive vocabulary than early gesture use. To explore this possibility, we used the 14-month MCDI language comprehension score, which was available on 7 of the 11 children (the parents of 2 of the 4 missing children did not return the MCDI; the other 2 children did not begin our study until 18 months and the age-appropriate version of the MCDI for 18-month-olds does not contain a language comprehension measure). As expected, we found that 14-month MCDI language comprehension scores were related to 14-month gesture production measures ($\rho = .75$, $p = .05$). However, 14-month MCDI language comprehension scores showed only a small, and not significant, correlation with 30-month PPVT scores ($\rho = .25$, $p = .59$). In contrast, 14-month gesture production for these 7 children showed a higher, and marginally significant, correlation with 30-month PPVT ($\rho = .63$, $p = .07$). Consistent with these findings, in analyses of the 53 TD children who served as the normative base for the current study, Rowe and Goldin-Meadow (2009b) found that 14-month language comprehension, as measured by the CDI, did *not* relate to children's 42-month PPVT scores, whereas 14-month gesture production did. Although early gesture and early language comprehension skills have *both* been found to predict later language production in previous work on late-talkers (Thal & Tobias, 1992), the power of the two predictors has not been directly compared. Our findings suggest that early gesture may be the more powerful predictor in children with PL and in TD children (Rowe & Goldin-Meadow, 2009b).

Lesion Characteristics and Gesture

We have found that early gesture use predicts later productive and receptive vocabulary in children with PL. Does gesture use vary with lesion characteristics? The small number of children in our sample of children with PL, combined with their diversity in lesion location, size, and type, makes it difficult to explore this question (see Rowe, Levine, Fisher, & Goldin-Meadow, 2009, for discussion). We therefore used nonparametric chi-square analyses to compare the distribution of children in the LOW vs. HIGH gesture groups as a function of each of the three lesion characteristics: lesion type (CV or PV), lesion size (small or large), and hemisphere (right or left). However, we stress that the findings must be considered tentative given our sample size of 11 children.

We found a significant effect for lesion type ($\chi^2 = 5.24$, $df = 1$, $p = .02$): All 4 children with PV lesions fell into the HIGH gesture group (none in the LOW gesture group), compared to only 2 of 7 children with CV infarcts (5 in the LOW gesture group). We found a trend for lesion size ($\chi^2 = 2.396$, $df = 1$, $p = .12$): One of the 5 children with small lesions fell into the LOW gesture group and 4 fell into the HIGH; in contrast, 4 of the 6 children with large lesions fell into the LOW gesture group and 2 fell into the HIGH. We did not find a significant result for lesion hemisphere ($\chi^2 = 0.749$, $df = 1$, $p = .39$): Three of the 8 children with left hemisphere lesions fell into the LOW gesture group and 5 fell into the HIGH, compared to 2 and 1, respectively, of the 3 children with right hemisphere lesions (but note that there were only 3 children in the right hemisphere lesion group).

Overall, children with small (as opposed to large) lesions and children with PV lesions (as opposed to CV infarcts) were more likely to produce gestures at a typical rate at 18 months. In turn, these children were also more likely to develop vocabulary at a typical rate than those with CV infarcts and those with large lesions, as evidenced by the significant relation between 18-month gesture and 22- and 26-month productive vocabulary and between 18-month gesture and 30-month receptive vocabulary. These findings are consistent with our previous studies including larger samples of children with PL (Rowe et al., 2009).

Discussion

Consistent with previous studies, we found that children with PL experience language delays that vary in terms of their severity and their persistence over time. Our study goes beyond previous studies of children with PL in that we observed gesture as well as language use in the children. We not only found great variability in gesture use at 18 months in children with PL, but we also found that this variability was related to later vocabulary levels—the children who produced fewer gesture types at 18 months were just the children who exhibited lower scores on productive and receptive vocabulary measures at 30 months. Most dramatically, child gesture at 18 months predicted child receptive vocabulary at 30 months, even when controlling for child speech types at 18 months.

Our findings extend previous research conducted on other child clinical groups, which report relations between gesture and language (Capone & McGregor, 2004). For example, Hill, Bishop, and Nimmo-Smith (1998) found gesture deficits in children with specific language impairment. As another example, Thal and Tobias (1992) observed communicative gesture use in 18- to 28-month-old late talkers, none of whom had been diagnosed with brain injury. The late talkers whose productive language had normalized at the 1-year follow-up had, at the earlier ages, produced significantly more communicative gestures than the late talkers whose productive language delays persisted at the 1-year follow-up. A tight relation between gesture and language has also been reported in adult Broca's and Wernicke's aphasics (Cicone, Wapner, Foldi, Zurif, & Gardner, 1979; McNeill, 1992). Taken together, these findings indicate that the relation between gesture and language is robust, found not only in samples of TD children but also in a variety of clinical populations.

Our findings have both theoretical and practical implications. In theoretical terms, the findings provide further evidence that the integrated gesture–language system found in TD children may be fundamental to the language learning process. Indeed, the diversity of our sample of children with PL underscores the robustness of this relation. Despite the fact that the children with PL varied widely in the location, size, and type of lesion they sustained, their early gestures were a reliable index of their later productive and receptive language skills. Although a larger sample is clearly needed to fully understand the relation between gesture and language, the fact that gesture and language remain linked in a population whose language functions are being carried out by a wider variety of neural substrates than is typical suggests that early gesture may be inextricably linked to the language learning process.

The theoretical question raised by our data is—why does gesture predict subsequent vocabulary development? Gesturing and word learning may draw on a common underlying cognitive skill, for example, the ability to produce and understand symbols (cf. Bates et al., 1979; Bretherton & Bates, 1984; McNeill, 1992; Werner & Kaplan, 1963). Gesturing and word learning may also be affected in similar ways by more general skills, including overall cognitive functioning (as measured by IQ tests, Levine et al., 1987) or general motor abilities (particularly given recent speculation about the importance of motor movements in language processing; e.g., Nishitani, Schürmann, Amunts, & Hari, 2005; Pulvermüller, 2005; Rizzolatti & Arbib, 1998). Indeed, previous findings have linked motor and language skills in TD individuals (Bates & Dick, 2002; Iverson & Thelen, 1999) and in clinical populations for whom some form of delay in language production is central (Bishop & Edmundson, 1987; Corriveau & Goswami, 2009; Hill, 2001; Visscher, Houwen, Scherder, Moolenaar, & Hartman, 2007). Individuals with PL, including many of the children in our sample, have hemiparesis of varying severity (Huttenlocher, 2002; Levine et al., 1987; Mercuri et al., 2004; Staudt et al., 2002; Staudt et al., 2004; Steenbergen & Gordon, 2006). Although it has not been shown that the severity of their symptoms varies systematically with their language skills, severity of hemiparesis has been shown to vary with IQ scores (see, e.g., Feldman, Janosky, Scher, & Wareham, 1994; Levine et al., 1987).

To summarize thus far, gesturing may predict subsequent vocabulary development because it is affected by cognitive or motor skills in the same way that word learning is. If so, gesture will be able to serve as a good index of vocabulary growth, but it will play no role in bringing that growth about. It is possible, however, that gesture goes beyond *reflecting* skill in vocabulary learning to actually play a role in *causing* vocabulary learning (Goldin-Meadow, 2003). Consider, for example, a child who points at a bird. Mother responds by saying, "Yes, that's a bird," thus providing a timely word-learning model for the child (see Goldin-Meadow, Goodrich, Sauer, & Iverson, 2007). In other words, gesture may be playing a role in vocabulary learning by allowing children to elicit input targeted to their level and, in this way, shape their own learning environments. Gesture could also play a role in vocabulary learning more directly by providing children with the opportunity to refer to objects whose names they are not yet able to express in speech, and to do so in a communicative context. Repeated use of gestures to refer to particular objects in an act of communication could pave the way for later acquisition of the words that refer to these particular objects (cf. Iverson & Goldin-Meadow, 2005). If so, gesture may be playing a role in the process of change by affecting learners themselves.

Our findings also have practical implications. The findings suggest that early gesture provides a way to identify children with persistent versus transient language difficulties at a time when all of the children are saying very little. Consistent with previous research with children with PL, 8 of the 11 children in our sample of children with PL fell below the 25th percentile of TD children on vocabulary production at 18 months, but only 5 children continued to be below the 25th percentile at 22 months. The striking result of our study is that these 5 were the same 5 children who were low gesture producers at 18 months. Thus, early gesture may provide clinicians with a way to identify children who may end up having persistent language difficulties *before* those difficulties appear in the children's speech. This early identification may then be useful in guiding intervention practices that may be particularly effective when employed early in development. Larger samples are clearly needed to explore gesture's use as a diagnostic indicator in clinical practice (e.g., to determine what the cutoff in gesture use ought to be to reliably distinguish children with transient vs. persistent language delays). However, our study suggests that early child gesture may be a useful indicator of subsequent language delay and thus motivates future research with both children with PL and children without PL who experience language delays.

Our results also raise the possibility that encouraging children to gesture may prove to be an effective intervention. That is, early gesture may not only precede early language but may facilitate its development. Studies of TD elementary school children have shown that the knowledge children convey uniquely in their gestures when explaining their solutions to mathematical equivalence problems is a harbinger of the knowledge that they will soon acquire (Alibali & Goldin-Meadow, 1993; Perry, Church, & Goldin-Meadow, 1988). More importantly, children who were told to gesture during a mathematical equivalence lesson retained more from the lesson than children who were told not to gesture (Cook, Mitchell, & Goldin-Meadow, 2008; see also Broaders, Cook, Mitchell, & Goldin-Meadow, 2007; Goldin-Meadow, Cook, & Mitchell, 2009), suggesting that the act of gesturing itself facilitates learning. Research on the language development of TD children suggests that gesture may play a similar role in language learning. For example, Goodwyn, Acredolo, and Brown (2000) found that encouraging children to gesture (via parent modeling) facilitates subsequent language development. Thus, gesture might serve not only as a tool for diagnosing language delay but also as a tool for early intervention.

In summary, we have found that the relation between early gesture and later language, reported in TD children, is also found in children with early unilateral brain injuries. The tight relation between early gesture and subsequent language development thus appears to be a robust aspect of language development. Our findings further suggest that early delays in gesture production can be used to identify those children with PL whose language learning is likely to be delayed in the future. If so, we may be able to offer these children interventions while their language-learning trajectory is most malleable.

References

Alibali, M., & Goldin-Meadow, S. (1993). Gesture-speech mismatch and mechanisms of learning: What the hands reveal about a child's state of mind. *Cognitive Psychology, 25*, 468–523.

Bates, E., Benigni, L., Bretherton, I., Camaioni, L., & Volterra, V. (1979). *The emergence of symbols*. New York: Academic Press.

Bates, E., & Dick, F. (2002). Language, gesture, and the developing brain. *Developmental Psychobiology, 40*, 293–310.

Bates, E., Thal, D., & Janowsky, J. (1992). Early language development and its neural correlates. In S. Segalowitz & I. Rapin (Eds.), *Handbook of neuropsychology: Vol. 7. Child neuropsychology* (pp. 69–110). Amsterdam: Elsevier.

Bates, E., Thal, D., Trauner, D., Fenson, J., Aram, D., Eisele, J., et al. (1997). From first words to grammar in children with focal brain injury. *Developmental Neuropsychology, 13*, 275–343.

Bishop, D. V. M., & Edmundson, A. (1987). Specific language impairment as a maturational lag—Evidence from longitudinal data on language and motor development. *Developmental Medicine and Child Neurology, 29*, 442–459.

Brasky, K., Nikolas, M., Meanwell, C., Levine, S. C., & Goldin-Meadow, S. (2005, June). *Language development in children with unilateral brain injury: Effects of lesions size.* Poster session presented at the Symposium on Research in Child Language Disorders, Madison, WI.

Bretherton, I., & Bates, E. (1984). The development of representation from 10 to 28 months. In R. Emde & R. Harmon (Eds.), *Continuities and discontinuities of development* (pp. 229–261). New York: Plenum.

Broaders, S., Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2007). Making children gesture brings out implicit knowledge and leads to learning. *Journal of Experimental Psychology: General, 136,* 539–550.

Capone, N., & McGregor, K. (2004). Gesture development: A review for clinical and research practices. *Journal of Speech, Language, and Hearing Research, 47,* 173–186.

Chilosi, A., Cipriani, P., Bertuccelli, B., Pfanner, L., & Cioni, G. (2001). Early cognitive and communication development in children with focal brain lesions. *Journal of Child Neurology, 16,* 309–316.

Cicone, M., Wapner, W., Foldi, N., Zurif, E., & Gardner, H. (1979). The relation between gesture and language in aphasic communication. *Brain and Language, 8,* 324–349.

Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition, 106,* 1047–1058.

Corriveau, K. H., & Goswami, U. (2009). Rhythmic motor entrainment in children with speech and language impairments: Tapping to the beat. *Cortex, 45,* 119–130.

Dunn, L., & Dunn, L. (1997). *Peabody Picture Vocabulary Test–Third Edition (PPVT–3).* Circle Pines, MN: American Guidance Service.

Feldman, H. M. (2005). Language learning with an injured brain. *Language Learning and Development, 1,* 265–288.

Feldman, H. M., Holland, A. L., Kemp, S. S., & Janosky, J. E. (1992). Language development after unilateral brain injury. *Brain and Language, 42,* 89–102.

Feldman, H. M., Janosky, J. E., Scher, M. S., & Wareham, N. L. (1994). Language abilities following prematurity, periventricular brain injury, and cerebral palsy. *Journal of Communication Disorders, 27,* 71–90.

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., et al. (1993). *The MacArthur Communicative Development Inventories: User's guide and technical manual.* Baltimore: Brookes.

Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think.* Cambridge, MA: Harvard University Press.

Goldin-Meadow, S., & Butcher, C. (2003). Pointing toward two-word speech in young children. In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 85–107). Hillsdale, NJ: Erlbaum.

Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. A. (2009). Gesturing gives children new ideas about math. *Psychological Science, 20,* 267–272.

Goldin-Meadow, S., Goodrich, W., Sauer, E., & Iverson, J. (2007). Young children use their hands to tell their mothers what to say. *Developmental Science, 10,* 778–785.

Goldin-Meadow, S., & Mylander, C. (1984). Gestural communication in deaf children: The effects and noneffects of parental input on early language development. *Monographs of the Society for Research in Child Development, 49* (Serial No. 207).

Goodwyn, S., Acredolo, L., & Brown, C. (2000). Impact of symbolic gesturing on early language development. *Journal of Nonverbal Behavior, 24,* 81–103.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children.* Baltimore: Brookes.

Hesterberg, T., Moore, D. S., Monaghan, S., Clipson, A., & Epstein, R. (2006). Bootstrap methods and permutation tests. In D. S. Moore & G. P. McCabe (Eds.), *Introduction to the practice of statistics* (5th ed., pp. 14–70). New York: Freeman.

Hill, E. L. (2001). Non-specific nature of specific language impairment: A review of the literature with regard to concomitant motor impairments. *International Journal of Language and Communication Disorders, 36,* 149–171.

Hill, E. L., Bishop, D. V. M., & Nimmo-Smith, I. (1998). Representational gestures in developmental coordination disorder and specific language impairment: Error-types and the reliability of ratings. *Human Movement Sciences, 17,* 655–678.

Huttenlocher, P. (2002). *Neural plasticity: The effects of environment on the development of the cerebral cortex.* Cambridge, MA: Harvard University Press.

Iverson, J., Capirci, O., & Caselli, M. C. (1994). From communication to language in two modalities. *Cognitive Development, 9,* 23–43.

Iverson, J., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science, 16,* 367–371.

Iverson, J. M., & Thelen, E. (1999). Hand, mouth, and brain: The dynamic emergence of speech and gesture. *Journal of Consciousness Studies, 6,* 19–40.

Krägeloh-Mann, I., & Horber, V. (2007). The role of magnetic resonance imaging in elucidating the pathogenesis of cerebral palsy: A systematic review. *Developmental Medicine & Child Neurology, 49,* 144–151.

Levine, S. C., Huttenlocher, P., Banich, M. T., & Duda, E. (1987). Factors affecting cognitive functioning of hemiplegic children. *Developmental Medicine and Child Neurology, 29,* 27–35.

Levine, S. C., Kraus, R., Alexander, E., Suriyakham, L., & Huttenlocher, P. (2005). IQ decline following early unilateral brain injury: A longitudinal study. *Brain and Cognition, 59,* 114–123.

Marchman, V., Miller, R., & Bates, E. (1991). Babble and first words in children with focal brain injury. *Applied Psycholinguistics, 12,* 1–22.

McNeill, D. (1992). *Hand and mind.* Chicago: University of Chicago Press.

Mercuri, E., Barnett, A., Rutherford, M., Guzzetta, A., Haataja, L., Cioni, G., et al. (2004). Neonatal cerebral infarction and neuromotor outcome at school age. *Pediatrics, 113*, 95–100.

Nishitani, N., Schürmann, M., Amunts, K., & Hari, R. (2005). Broca's region: From action to language. *Physiology, 20*, 60–69.

O'Reilly, A. W., Painter, K. M., & Bornstein, M. H. (1997). Relations between language and symbolic gesture development in early childhood. *Cognitive Development, 12*, 185–197.

Özçaliskan, S., & Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language development. *Cognition, 96*, B101–B113.

Perry, M., Church, R. B., & Goldin-Meadow, S. (1988). Transitional knowledge in the acquisition of concepts. *Cognitive Development, 3*, 359–400.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience, 6*, 576–582.

Reilly, J., Levine, S. C., Nass, R., & Stiles, J. (2008). Brain plasticity: Evidence from children with perinatal brain injury. In J. Reed & J. Warner (Eds.), *Child neuropsychology* (pp. 58–91). Oxford, UK: Blackwell.

Rizzolatti, G., & Arbib, M. (1998). Language within our grasp. *Trends in Neurosciences, 21*, 188–194.

Rowe, M., & Goldin-Meadow, S. (2009a). Differences in early gesture explain SES disparities in child vocabulary size at school entry. *Science, 323*, 951–953.

Rowe, M., & Goldin-Meadow, S. (2009b). Early gesture selectively predicts later language learning. *Developmental Science, 12*, 182–187.

Rowe, M. L., Levine, S. C., Fisher, J., & Goldin-Meadow, S. (2009). Does linguistic input play the same role in language learning for children with and without early brain injury? *Developmental Psychology, 45*, 90–102.

Rowe, M., Özçaliskan, S., & Goldin-Meadow, S. (2006). The added value of gesture in predicting vocabulary growth. In D. Bamman, T. Magnitskaia, & C. Zaller (Eds.), *Proceedings of the 30th Annual Boston University Conference on Language Development* (pp. 501–512). Somerville, MA: Cascadilla Press.

Rowe, M., Özçaliskan, S., & Goldin-Meadow, S. (2008). Learning words by hand: Gesture's role in predicting vocabulary development. *First Language, 28*, 185–203.

Slobin, D. I., & Welsh, C. A. (1967). *Elicited imitation as a research tool in developmental psycholinguistics* (ERIC Document Reproduction Service No. ED012892).

Staudt, M., Gerloff, C., Grodd, W., Holthausen, H., Niemann, G., & Krägeloh-Mann, I. (2004). Reorganization in congenital hemiparesis acquired at different gestational ages. *Annals of Neurology, 56*, 854–863.

Staudt, M., Grodd, W., Gerloff, C., Erb, M., Stitz, J., & Krägeloh-Mann, I. (2002). Two types of ipsilateral reorganization in congenital hemiparesis: A TMS and fMRI study. *Brain, 125*, 2222–2237.

Steenbergen, B., & Gordon, A. (2006). Activity limitation in hemiplegic cerebral palsy: Evidence for disorders in motor planning. *Developmental Medicine & Child Neurology, 48*, 780–783.

Stiles, J., Reilly, J., Paul, B., & Moses, P. (2005). Cognitive development following early brain injury: Evidence for neural adaptation. *Trends in Cognitive Sciences, 9*, 136–143.

Thal, D., Marchman, V., Stiles, J., Aram, D., Trauner, D., Nass, R., et al. (1991). Early lexical development in children with focal brain injury. *Brain and Language, 40*, 491–527.

Thal, D., & Tobias, S. (1992). Communicative gestures in children with delayed onset of oral expressive vocabulary. *Journal of Speech and Hearing Research, 35*, 1281–1289.

Vicari, S., Albertoni, A., Chilosi, A., Cipriani, P., Cioni, G., & Bates, E. (2000). Plasticity and reorganization during language development in children with early brain injury. *Cortex, 36*, 31–46.

Visscher, C., Houwen, S., Scherder, E. J. A., Moolenaar, B., & Hartman, E. (2007). Motor profile of children with developmental speech and language disorders. *Pediatrics, 120*, E158–E163.

Werner, H., & Kaplan, B. (1963). *Symbol formation: An organismic-developmental approach to language and the expression of thought*. New York: Wiley.

Woods, B., & Teuber, H. (1978). Changing patterns of childhood aphasia. *Annals of Neurology, 3*, 273–280.

Child Development, March/April 2010, Volume 81, Number 2, Pages 540–554

Mother–Infant Interaction During the First 3 Months: The Emergence of Culture-Specific Contingency Patterns

Joscha Kärtner, Heidi Keller, and Relindis D. Yovsi
University of Osnabrück

This study analyzed German and Nso mothers' auditory, proximal, and visual contingent responses to their infants' nondistress vocalizations in postnatal Weeks 4, 6, 8, 10, and 12. Visual contingency scores increased whereas proximal contingency scores decreased over time for the independent (German urban middle-class, $N = 20$) but not the interdependent sociocultural context (rural Nso farmers, $N = 24$). It seems, therefore, that culture-specific differences in the modal patterns of contingent responsiveness emerge during the 2nd and 3rd months of life. This differential development was interpreted as the result of the interplay between maturational processes associated with the 2-month shift that are selectively integrated and reinforced in culture-specific mother–infant interaction.

Caretakers' contingent responsiveness to their infants' communicative signals is a key feature of early parenting. Mothers intuitively react to infants' signals within a short period of time, normally less than a second (Keller, Lohaus, Völker, Cappenberg, & Chasiotis, 1999; Papoušek & Papoušek, 1987). Infants, in turn, are motivated to detect these contingencies; this behavior is self-rewarding in that it gives rise to positive affect (Watson, 1985). In detecting these contingencies, infants begin to perceive themselves as causal agents whose behavior directly affects their social environment. This knowledge has fundamental implications for infants' subsequent cognitive and socioemotional developmental processes (cf. Bigelow, 1998; Keller, Kärtner, Borke, Yovsi, & Kleis, 2005; Tarabulsy, Tessier, & Kappas, 1996).

Contingent responsiveness can be regarded as a universal, evolutionarily shaped component of parenting that is realized in context-sensitive ways. Support for this assumption comes from studies that explored the universality of contingent responsiveness by looking at cross-cultural similarities and differences in caretakers' responses. The general consensus is that the overall level of contingent responsiveness is similar across cultural contexts; however, there are differences in terms of the way in which caretakers react contingently (Bornstein et al., 1992; Fogel, Toda, & Kawai, 1988; Kärtner et al., 2008; Richman, Miller, & LeVine, 1992). For example, Richman et al. (1992) found that Gusii mothers from Western Kenya were predominantly physically responsive while Euro-American middle-class mothers were predominantly verbally and visually responsive. In a similar study, Kärtner et al. (2008) compared the contingent responses of caretakers from independent, autonomous-relational, and interdependent sociocultural contexts. Kärtner et al. found that caretakers from independent sociocultural contexts addressed the infants' sense of sight significantly more often than did caretakers from autonomous-relational or interdependent contexts. Infants from interdependent contexts, on the other hand, experienced significantly more proximal contingent responses addressing the vestibular sense or the sense of touch. Kärtner et al. argued that it was the difference in the caretakers' broader sociocultural orientation and cultural embedment that led to these modal differences in contingent responsiveness. Other researchers have offered similar explanations, such as different norms of interaction (Richman et al., 1992) or differences in cultural beliefs (Fogel et al., 1988), to

---

We thank Arnold Lohaus for his cooperation in assessing the Münster data as part of a joint research project funded by the German Research Foundation (LO 337/12–1, KE 263/17–4). We are grateful and indebted to the German Research Foundation for supporting the assessment and analysis of the Nso data (KE263/41). Many thanks go to our graduate students Ute Spilker, Christina Heinrichs, Vanessa Tepe, Antje Beck, and Carmen Reinert. We are especially grateful to the infants and caretakers who participated in this study.

Correspondence concerning this article should be addressed to Joscha Kärtner, Faculty of Human Sciences, Culture, Learning and Development, University of Osnabrück, 49069 Osnabrück, Germany. Electronic mail may be sent to joscha.kaertner@uni-osnabrueck.de.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8102-0011

account for cross-cultural variation in contingent responsiveness.

There is some empirical evidence for a functional link between the dominant sociocultural orientation, actual parenting behavior, and child development. Keller, Yovsi, et al. (2004), for example, have shown that culturally informed caretaking behavior fosters desirable developmental outcomes. Specifically, Keller, Yovsi, et al. found that proximal caretaking, mainly in the form of body contact, during mother–infant interaction when infants were 3 months old led to earlier development of compliance and obedience in the toddlers' 2nd year. Both proximal caretaking and obedience of toddlers are highly valued in interdependent sociocultural contexts (Keller, 2007; Ogunnaike & Houser, 2002; Whiting & Pope Edwards, 1988). In contrast to proximal caretaking, distal parenting, mainly in the form of object stimulation and mutual gaze, led to earlier development of the toddlers' autonomous self-concept as assessed by mirror self-recognition. Both distal parenting and toddlers' autonomy are highly valued in predominantly independent sociocultural contexts (Keller, 2007; LeVine, 1994; Richman et al., 1992).

As described above, parents from interdependent and independent sociocultural contexts differ in their contingency patterns and, more generally, in their parenting behavior. For example, German middle-class families have a predominantly independent sociocultural orientation, whereas rural Nso families have a primarily interdependent sociocultural orientation (cf. Kärtner et al., 2008; Keller, 2007; Yovsi, 2003). The Nso, an ethnic group living in a mainly subsistence-based agrarian ecology in Cameroon, socialize their children toward obedience, respect for elders, cooperation, and social responsibility (Nsamenang, 1992). Obligation between kin people is an integral and institutionalized part of Nso morality and daily life (Goheen, 1996) and parents foster their children's integration into a hierarchically structured social setting (Nsamenang & Lamb, 1994).

German middle-class mothers, in contrast, socialize their children toward individuality, autonomy (Keller, 2007), and self-reliance (LeVine & Norman, 2001). Mothers value exclusive dyadic interaction and believe that children need to spend time on their own in order to become more independent (Keller, Hentschel, et al., 2004; Keller, Voelker, & Yovsi, 2005). Based on these differences in sociocultural orientation, we hypothesized that in the 3rd month of life, Nso infants would experience more proximal contingent responses addressing the sense of touch and the vestibular sense and less visual contingencies than would infants from the German middle-class sample. Across modalities, however, mothers from the two sociocultural contexts should show approximately the same amount of contingent responses to their infants' signals; that is, only the modality patterns of the contingent responses should differ between the groups.

The First 3 Months: Continuity of Culture-Specific Patterns or Differential Developmental Change?

Given that the majority of studies of contingent responsiveness focuses on infants who are at least 3 months old, it is as yet unknown whether culture-specific manifestations are present from birth or whether different patterns emerge across the infants' first few months of life. In this way, culture-specific contingency patterns could either be conceptualized as (a) a continuous behavioral response informed by broad and static cultural scripts that guide mothers' behavior or (b) the outcome of a more dynamic, culture-specific interplay of the infants' developing communicative competencies and maternal responses.

Research on the early development of contingent responsiveness is scarce and offers mixed results, partly in support of the continuity hypothesis and partly in support of the developmental change hypothesis. With respect to continuity, the assumption that mothers' contingent responsiveness is a relatively stable feature of mother–infant interaction was first discussed by Watson (1985) and Stern (1985). Both authors argued that social interaction creates default contingency levels through familiarization and that these contingency levels are subsequently reflected in infants' responses to new social stimuli. More recently, Bigelow and colleagues provided empirical evidence in support of the notion of continuity of contingent responsiveness (Bigelow, 1998; Bigelow & DeCoste, 2003; Bigelow & Rochat, 2006). The authors found that 2-month-old infants were most responsive to strangers who showed similar levels of contingent responsiveness as the infants' own mothers (Bigelow & Rochat, 2006). These results indicate that infants expected specific contingency levels, which implies that the mothers' behavior had been sufficiently continuous as to build up these expectations in their infants. Further support for the continuity hypothesis comes from Hsu and Fogel (2003), who found that from postnatal Week 4 to Week 24 there was no effect of age on maternal behavioral actions (verbal/vocal, facial expression, touch, and head movement) in response to infants' nondistress vocalizations (NDVs). Taken together, these results suggest that, at least over short periods of time, mothers' responsiveness is reasonably continuous.

Evidence for developmental change in contingent responsiveness comes from research examining the 2-month shift. The 2-month shift refers to the period near the end of an infant's 2nd month of life wherein there is a developmental change that marks a qualitative shift in the way in which the infant interacts with social partners and expresses himself or herself (Emde, 1984). Following this shift, infants spend more time in an alert active state (Wolff, 1987), their ability to maintain visual attention increases (Haith, Bergman, & Moore, 1977), and they start smiling socially (Wolff, 1987). There have been several studies that have demonstrated that these changes in infants affect mothers' responsiveness. First, Lavelli and Fogel (2005) found that mothers reacted contingently by talking and smiling in response to their infants' smiles and cooing more during the 2nd and 3rd months than in the 1st month. Second, Symons and Moran (1994) found that there was a steady increase in mothers' contingent smiles in response to infants' smiles from the 2nd to the 6th month. This result is inconclusive, however, as the authors did not report inferential test statistics for the increase. Last, Lavelli and Fogel (2002) have shown that interdyadic differences in face-to-face communication, especially in active engagement between mother and infant, were most likely to emerge during the crucial developmental transition period, that is, toward the end of the 2nd month. Based on these findings, we hypothesized that if mothers' contingency patterns change across the first 3 months, then these changes should coincide with the occurrence of the 2-month shift as indexed by the infants' awake alertness and gazing pattern (especially looks directed at their mothers).

The findings on developmental change in contingent responsiveness, as discussed above, were mostly related to distal aspects of mother–infant interaction, that is, smiling and gazing patterns. These patterns are of central importance in predominantly independent sociocultural contexts. Cross-cultural research has demonstrated that face-to-face interaction does not have the same prevalence and may not have the same meaning across different sociocultural contexts (Keller, 2007). This difference in prevalence and meaning across contexts may have significant implications for the mother's and especially the infant's behavior during early face-to-face interaction. For example, Lavelli and Fogel (2005) showed that mothers' expressions amplified and organized infants' patterns of attention and emotion in relatively stable sequences of mother–infant face-to-face communication in the 3rd month. If face-to-face interaction is less prevalent, however, the development of these stable sequences may be more characteristic of dyads from predominantly independent sociocultural contexts. The question remains as to whether the 2-month shift takes the same form in the interdependent sociocultural context, that is, whether there is a comparable increase in the infants' awake alertness and looking at their mothers. What are the implications of this result for mothers' contingent responsiveness? We addressed these two issues in the present study.

The overarching goal of the present study was to examine the development of cross-cultural modal differences in caretakers' patterns of contingent responsiveness. To do this, we examined the auditory, proximal, and visual contingent reactions that mothers from two different sociocultural contexts (independent and interdependent) show in response to their infants' NDVs in postnatal Weeks 4, 6, 8, 10, and 12. There are three specific hypotheses and two exploratory research questions. First, the overall level of contingent responsiveness should be continuous across time and should not differ between sociocultural contexts. Second, in the 3rd month, infants from the independent sociocultural context should experience more visual contingencies and less proximal contingencies than should infants from the interdependent sociocultural context. Third, contingency patterns should reflect changes associated with the 2-month shift. More specifically, the proportion of mothers' visual contingencies should increase with changes in their infants' awake alertness and gazing patterns. Furthermore, we explored whether the 2-month shift in awake alertness and infants' gazing occurred to the same degree for infants from both sociocultural contexts and whether the consequences for their mothers' contingent responsiveness were similar for the two groups. Finally, we analyzed the stability of contingent responsiveness, the provision of face-to-face context, and the infants' gazing at their mothers' faces across the infants' 2nd and 3rd months in order to learn more about the behavioral organization of early mother–infant interaction. Since little is known about this issue, we acted on the assumption that stabilities should be reasonably high in both sociocultural contexts.

In order to keep the procedure for assessing mother–infant interactions ecologically valid, mothers were not constrained to a face-to-face interaction but were free to choose a typical setting for interacting with their infants. Rochat (2001) and others (e.g., Jaffe, Beebe, Stanley, Crown, & Jasnow, 2001; Watson, 1985) argue that contingency levels are embedded in face-to-face interactions during mother–infant routines. Following this logic, we would expect infants from interdependent contexts to experience less contingent responses. Kärtner et al. (2008), however, argue that caretakers' contingent responsiveness does not presuppose face-to-face interaction. The results of their study showed that, independent of differences in the extent of caretakers' provision of face-to-face interaction, the overall contingency levels across distal (vocal/verbal, smile, facial expression) and proximal (body contact, body stimulation) modalities were similar across sociocultural contexts.

Regarding the type of infant signals that mothers respond to, a number of studies have shown that, compared to several nonverbal behaviors, infants' NDVs constitute the most salient signal for caretakers (Bornstein et al., 1992; Hsu & Fogel, 2003; Van Egeren, Barratt, & Roach, 2001). This is a consistent finding across sociocultural contexts as diverse as a rural Nso community living in the Bantu grass fields in the North Western province in Cameroon or urban middle-class families from Berlin or Los Angeles (Kärtner et al., 2008).

Method

Participants

Twenty families from Münster, a medium-sized city in Northern Germany, and 24 Nso families living in small villages around Kumbo, Cameroon, participated in the present study. In Germany, the mean age of first-time mothers was approximately 29 years and the fertility rate was 1.4 children born per woman in the last 10 years (Statistisches Bundesamt, 2008). In German middle-class families, infants spend most of their time at home with their mothers during the 1st month. Older siblings go to day-care facilities or nursery school in the mornings. The role distribution in most German families is still traditional in that it is mostly the mothers who take care of the infants while the fathers are at work (Keller, 2006; Keller, Zach, & Abels, 2005).

The Nso sample was recruited from Kikaikelaki, a village near Kumbo, Bui Division of the Northwest Province of Cameroon. In 2004, the average life expectancy in Cameroon was 48.0 years and the fertility rate was 4.6 children born per woman (Central Intelligence Agency, 2008). The mean age of first-time Nso mothers was approximately 19.8 years in 2003 (Yovsi, 2003). The majority of the Nso population live in village communities where subsistence is a combination of communal efforts and endeavors from all family members, including children (Yovsi, 2003). Men achieve high social status through the possession of *keng* (wealth) in the form of raffia palms, coffee plantations, forests, and houses. In Kikaikelaki there are denominational primary schools and most children end their education at the elementary level (Class 7) at, on average, 13–14 years. There are health centers in the villages but they lack infrastructure and most people prefer traditional medicine. Nso villages are made up of several compounds, consisting of houses grouped around a central yard. The settlement pattern is patrilocal and every male owns land within the lineage territory where he builds and settles with his family. The rural Nso have an extended family system of three or more generations. Families share everyday activities, including child care. Thus, children are socialized in a dense social network including parents, siblings, relatives, grandparents, and neighbors. Child care is a communal responsibility; after the birth of a child, everybody in the community has an obligation toward his or her care and social development (Nsamenang, 1992; Yovsi, 2003). Although the German middle-class and the rural Nso families differ in terms of the size of the social networks in which children are socialized, in both of the sociocultural groups the mother is the primary caregiver during the first 6 months of life (Yovsi, 2003).

All of the mothers were first contacted during the last trimester of pregnancy through a pediatric hospital in Münster and through the health centre in Kumbo. In the rural Nso community, mothers were informed only after the lineage heads approved participation. Interested mothers participated if their family heads consented.

There were 50 families in the original sample, however, six Nso families were excluded for equipment or procedural error ($N = 4$) or because two or more observations were lacking ($N = 2$), leaving a total sample of 44 families. Infant gender was balanced across samples (Münster: 45% girls, Nso: 54% girls). The Nso mothers were slightly younger ($M = 27.40$, $SD = 8.15$ years) than were the mothers from Münster ($M = 30.70$, $SD = 3.76$ years), $t(42) = 1.67$, $p > .10$, $d = 0.51$. All of the mothers from Münster and 33% of the Nso mothers were first-time mothers, $\chi^2 = 20.95$, $p < .01$. The Nso mothers had received less formal education ($M = 8.00$, *SD* = 1.91 years) than had the mothers from Münster (*M* = 14.35, *SD* = 3.23 years), $t(42) = 8.09$, $p < .01$, $d = 2.45$. According to our sociostructural approach to culture, these patterns of sociodemographic differences are constitutive of the different sociocultural environments (Keller, 2007).

Procedure

After the infants' birth, local research assistants made appointments with and visited mothers at home when the infants were 4, 6, 8, 10, and 12 weeks of age (± 2 days). After the assistant had explained what the study entailed, the mothers, who were the primary caretakers in all instances, were instructed in their native language (German or Lamnso) to interact with their infants as they would normally do. The mothers were free to choose a typical setting. They were, however, asked not to move around too much and not to include others in the interaction. The interaction was videotaped for 10 min (*M* = 9 min 24 s, *SD* = 48 s) using a hand-held camera with a built-in microphone at a distance of approximately 2–3 m. The research assistants made sure that the infants were awake, fed, and not crying before filming began. Assessments with the Nso mother–infant dyads were mostly done outside the house, which is reflective of their daily family routines. A questionnaire concerning sociodemographic information was also administered during the first visit.

Coding

In order to analyze the caretakers' levels and patterns of contingent responsiveness we applied an event-coding approach. In Step 1, a pair of trained students identified the onset and offset of infant sounds for the whole 10 min. A new vocalization was coded when there was a pause of more than 1 s. The identified sounds were then evaluated according to type, intensity, and, in the case of vocalizations, valence. Three types of infant sounds were differentiated: (a) vocalizations, defined as all voiced articulations; (b) vegetative sounds such as sneezes, coughs, and hiccups; and (c) effort sounds, defined as sounds that occur in combination with movements or other efforts such as grunts or motor effort sounds. The intensity of infant sounds was coded as belonging to one of three categories: (a) low, defined as sounds that were uttered with low intensity, that is, hardly audible; (b) high, defined as sounds that were subjectively perceived as loud; or (c) medium, defined as sounds that were neither low nor high in intensity. The valence of vocalizations was judged as either: (a) negative if it was subjectively perceived as negative in hedonistic tone, such as whining, fussing, crying, sighing, and other sounds of discomfort; (b) positive if it was perceived as positive in hedonistic tone, such as sounds of happiness or laughing; or (c) neutral if it was neither negative nor positive, such as a-sounds or babbling.

With respect to interrater reliability, 21% of the sample was coded by both coders. For infant sounds, the proportion of agreement for onsets relative to the total number of onsets coded by at least one of the coders, within a time tolerance of 1 s between coders, was .77. Reliabilities were medium to high for the type ($\kappa = .79$), valence ($\kappa = .79$), and intensity ($\kappa = .81$) of the infants' sounds.

In the main analysis, we focused on the infants' NDVs, defined as neutral vocalizations of low or middle intensity (cf. Hsu & Fogel, 2003; Hsu, Fogel, & Messinger, 2001). On average, infants uttered 30.3 NDVs (*SD* = 21.6) per 10 min, which had a mean duration of 1.44 s each (*SD* = 3.30) and accounted for 55% of all infant sounds. These values were similar across cultural context and age. All other sounds were excluded for theoretical and methodological reasons. Theoretically, vegetative (13.0% of all infant sounds) and effort (19.4% of all infant sounds) sounds cannot be attributed to the infant's initiative; contingent responsiveness concerns the infant perceiving himself or herself as a causal agent. Furthermore, preliminary analyses indicated that caretakers reacted differently to negative (7.4% of all vocalizations), positive (4.9% of all vocalizations), and highly intense (4.5% of all vocalizations) vocalizations. As these vocalizations were distributed unequally across dyads, ages, and cultural contexts, differences in contingency patterns between participants or cultural groups would, in part, be caused by differential distributions of valence and intensity.

Based on the fact that maternal responses are most likely to occur in the 1st second after the infants' signal (Bloom, Russell, & Wassenberg, 1987; Keller et al., 1999; Papoušek & Papoušek, 1987), we defined a latency window of 1 s for contingency (cf. Bigelow, 1998; Bigelow & Rochat, 2006; Hsu & Fogel, 2003; Kärtner et al., 2008; Keller, Yovsi, et al., 2004; Keller et al., 2008). In other words, the caretakers' response had to occur within the latency window that begins with the onset and ends 1 s after the offset of the vocalization to be considered contingent.

In Step 2, the coders coded the caretakers' contingent responses. A caretaker's action was coded as contingent if the caretaker either initiated an action that was not shown before (*addition*) or changed the rhythm or the intensity of an ongoing action (*modification*; cf. Hsu & Fogel, 2003; Keller & Schölmerich, 1987). All other cases were coded as *no reaction*. The coding was carried out separately for the different senses that the caretakers' contingent response might have addressed: (a) *auditory contingencies* addressing the infants' sense of hearing were provided by the caretakers' verbalization, vocalization, and other sounds, for example, tongue clicking; (b) *proximal contingencies* addressing the sense of touch and the vestibular sense were provided by body stimulation defined as touching or stimulating the infant with the hands (e.g., stroking or patting) or the face (e.g., kissing) or by body contact; or (c) *visual contingencies* addressing the sense of sight were provided by the caretakers' gaze, smile, or facial expression (e.g., raised eyebrows and/or mouth wide open) or by using an object (e.g., a toy).

The 1 s before the onset of each NDV served as a baseline against which the caretakers' behavior was evaluated. In all cases, auditory, proximal, and visual contingencies were coded independently. For example, a mother might modify her vocal/verbal behavior and start smiling in response to an infant's NDV. Interrater reliabilities were computed on the basis of the same 21% that were coded by both coders. Cohen's kappas were medium to high for the modalities of the caretakers' contingent responses (body stimulation: $\kappa = .87$, body contact: $\kappa = .84$, gaze: $\kappa = .76$, smile: $\kappa = .85$, facial expression: $\kappa = .72$, vocalization: $\kappa = .89$, and object stimulation: $\kappa = .92$).

Approximately 4.5% of the NDVs were excluded from further analyses because not all of the caretakers' modalities could be coded (e.g., face covered by hair). In the final analysis, we computed a proportional score that indicated the *overall* contingency experience for each infant, that is, the percentage of NDVs that were responded to contingently by addressing at least one of the infant's senses. To analyze the similarities and differences in the patterns of contingent responses across the infants' age and sociocultural context, we computed three proportional scores that indicated the percentage of all contingent responses that were *auditory*, *proximal*, or *visual*.

In Step 3, a second pair of coders coded face-to-face context and the infants' alertness and gaze. In order to do this, we used an interval coding approach based on 10-s intervals. For each 10-s interval, the coders decided whether or not mothers created a context that was favorable for face-to-face interaction for at least half of the interval. A favorable context was defined as one in which the mother turned toward the infant and positioned herself and the infant in such a way that the mother's face was situated not more than 45° off her infant's optical axis. In addition, two indicators of the 2-month shift were coded: (a) the relative number of 10-s intervals in which the infant was alert, that is, when the infant did not show any signs of sleepiness (repeated yawning and/or drooping eyelids), and (b) the infants' gazing at their mothers' faces. Onset and offset of infants' gazing at their mothers' faces were coded during episodes where mothers had established face-to-face context and infants were alert. Interrater reliabilities that were computed on the basis of 21% of the data were medium to high for face-to-face context ($\kappa = .86$), alertness ($\kappa = .72$), and infants' gaze ($\kappa = .78$). The final scores for *face-to-face context* and *alertness* were the relative frequencies (i.e., proportions) of 10-s intervals in which the coding criteria applied so that the values represent the proportion of intervals during which the mother provided a favorable face-to-face context or the infant was awake, respectively.

For the infants' gazing at their mothers' face, we computed two scores. The first score, mutual gaze$_{\text{fav}}$, was defined as the duration of infants' gazing at their mothers divided by the absolute frequency of 10-s intervals in which mothers established face-to-face context *and* infants were alert multiplied by 10 s. The latter multiplication is an approximation of the total duration of context favorable for mutual gaze, that is, episodes where mothers had established face-to-face context and infants were alert. This score (mutual gaze$_{\text{fav}}$) indicates the duration of mutual gaze relative to the duration of episodes in which mothers established face-to-face context and infants were alert. The second score, mutual gaze, is the duration of infants' gazing at their mothers divided by the duration of the full interaction episode. Because infants' gaze was only coded during sequences where mothers established face-to-face context, this score approximates the duration of mutual gaze relative to the duration of the entire episode.

Results

Plan of Analysis and Missing Data Treatment

In order to analyze differences and similarities in maternal contingency scores, the provision of face-to-face context by caretakers, and the infants' gazing at their mothers and alertness across sociocultural contexts and age, we computed repeated measures analyses of variance (ANOVAs) with age (4, 6, 8, 10, 12 weeks) as the within-subjects factor and cultural group (Münster, Nso) as the between-subjects factor. To analyze the time course of age-related changes, we have reported linear and curvilinear contrasts by age and Age × Cultural Group interactions for each analysis. Significant interactions between age and sociocultural context were clarified by subjecting the age data to separate one-way ANOVAs for each cultural group. Post hoc *t* tests were performed in the case of significant main effects of age. As neither gender nor birth rank (first- vs. later-born) was significant in any of the preliminary analyses, the data were collapsed across these factors for all subsequent analyses. In order to evaluate the stability of contingent responsiveness, the provision of face-to-face context, and indicators of the 2-month shift across ages, we averaged correlations of adjacent time points (Weeks 4 and 6, Weeks 6 and 8, Weeks 8 and 10, and Weeks 10 and 12) separately for each cultural group. To compute averaged correlations, Pearson correlations were Fisher *z*-transformed, averaged, and then back-transformed.

In the final sample, the contingency data of 15 observations (5.7% of the sample) from 12 dyads (4 from Münster) were substituted by the values for adjacent time points of those dyads (van der Kloot, 1998). There were two reasons for substituting data. First, for 5 of the 44 families (3 from Münster) only four of the five observations took place. All of the other families were visited five times as planned. Second, some observations ($N = 10$; 1 from Münster) were excluded because there were only a few NDVs. Therefore, we used an inclusion criterion of at least three NDVs. Missing values of face-to-face ($N = 5$), mutual gaze ($N = 5$), and mutual gaze_{fav} ($N = 27$) were substituted according to the same logic. There were 22 observations (10%) for which mutual gaze_{fav} could not be computed because there was no favorable context which would be tantamount to a division by zero. To test the robustness of the results, all analyses were additionally conducted on the unsubstituted data, and a data set where missing values were substituted by the individual mothers' mean score. The pattern of significant effects and the magnitudes of the effects were identical across all three analyses. It should be noted here that, compared to the ANOVAs, in all of the correlational analyses that follow, only the data gathered during actual observations, that is, unsubstituted data, were analyzed.

Overall Levels and Patterns of Contingent Responsiveness

On average, mothers responded contingently by addressing at least one of the infant's senses to about 57.5% of all NDVs; this figure did not differ between cultural groups or age groups, largest $F(1, 42) = 2.76$, $p > .10$. There was a marginally significant linear contrast for the interaction between cultural group and age, indicating that there were differences in the linear development of overall contingency across age between the cultural groups, $F(1, 42) = 3.62$, $p < .10$, $\eta_p^2 = .08$. Looking at the two cultural groups separately, separate one-way ANOVAs yielded a marginally significant main effect of age on overall contingency for the Münster sample, $F(4, 76) = 2.42$, $p < .10$, $\eta_p^2 = .11$, but no systematic change with age. For the Nso sample, there was no main effect of age on overall contingency, $F(4, 92) = 1.35$, $p > .10$, $\eta_p^2 = .06$, and a marginally significant linear decrease with age, $F(1, 23) = 3.09$, $p < .10$, $\eta_p^2 = .12$ (see Table 1).

The scores for auditory, proximal, and visual contingencies were subjected to separate mixed ANOVAs. On average, 47.2% of the mother's contingent responses to the infant's NDVs were auditory, 43.8% were proximal, and 9% were visual. Auditory and proximal contingencies were independent of cultural context and infants' age, largest $F(1, 42) = 1.28$, $p > .10$. The marginally significant linear contrast of the interaction indicated that there were differences in the linear development of the proximal contingencies with age, $F(1, 42) = 2.85$, $p < .10$, $\eta_p^2 = .06$. Looking at the two cultural groups separately, one-way (age) ANOVAs yielded a significant main effect of age on proximal contingencies for the Münster sample, $F(4, 76) = 2.63$, $p < .05$, $\eta_p^2 = .12$, that could best be explained by a linear decrease with age, $F(1, 19) = 7.15$, $p < .05$, $\eta_p^2 = .27$. There were neither differences nor any systematic changes with age in the Nso sample (see Table 1). Post hoc *t* tests indicate that Nso infants experienced significantly more proximal contingencies than did infants from Münster in Week 12, $t(42) = -2.56$, $p = .014$, $d = -0.77$.

The largest differences pertained to visual contingencies. Infants from the Münster sample experienced more visual contingencies ($M = 13.00\%$, $SE = 1.87\%$) than did infants from the Nso sample ($M = 5.80\%$, $SE = 1.71\%$), $F(1, 42) = 8.07$, $p < .001$, $\eta_p^2 = .16$. On average, these visual contingencies

Table 1

Contingency Levels (Percentage) Across Time and Cultural Context

	Week 4	Week 6	Week 8	Week 10	Week 12
Overall					
Münster	62.56 (19.06)	52.79 (17.04)	64.32 (15.15)	62.57 (13.18)	62.45 (16.56)
Nso	58.33 (17.97)	55.06 (16.67)	56.85 (14.82)	52.73 (19.57)	50.17 (21.74)
Auditory					
Münster	48.56 (16.17)	41.20 (20.32)	45.96 (11.63)	42.66 (9.48)	49.50 (13.73)
Nso	48.85 (16.67)	45.38 (19.51)	51.36 (15.91)	51.41 (25.49)	45.41 (18.34)
Proximal					
Münster	43.80 (17.85)	48.83 (21.54)	41.24 (18.89)	40.45 (22.88)	32.80 (14.01)
Nso	45.55 (18.60)	47.12 (20.76)	44.90 (16.76)	43.53 (26.11)	47.50 (22.27)
Visual					
Münster	7.64 (8.88)	9.97 (13.84)	12.80 (14.27)	16.88 (17.04)	17.70 (13.52)
Nso	5.60 (8.46)	7.50 (8.88)	3.74 (7.67)	5.06 (7.49)	7.09 (12.28)

Note. Values are M (SD). $N = 20$ for the Münster and $N = 24$ for the Nso sample. Scores are either percentage of nondistress vocalizations (overall contingency level) or percentage of mothers' contingent responses (auditory, proximal, and visual contingency). Auditory, proximal, and visual contingencies add up to 100%.

increased with age, $F(4, 168) = 2.98$, $p < .05$, $\eta_p^2 = .07$. There was, however, a significant interaction, indicating that visual contingencies changed differently with age for the two cultural groups, $F(4, 168) = 3.04$, $p < .05$, $\eta_p^2 = .07$. Looking at the two cultural groups separately, one-way ANOVAs yielded a significant main effect of age for the Münster sample, $F(4, 76) = 3.08$, $p < .05$, $\eta_p^2 = .14$, that could best be explained by a linear increase with age, $F(1, 19) = 11.96$, $p < .01$, $\eta_p^2 = .39$. In contrast, there were neither differences nor any systematic changes with age for the Nso sample. Post hoc t tests yielded significant differences between the two cultural groups from Week 8 onward, $t_{\text{week 8}}(42) = 2.55$, $p < .017$, $d = 0.77$; $t_{\text{week 10}}(42) = 2.88$, $p < .01$, $d = 0.87$; and $t_{\text{week 12}}(42) = 2.73$, $p < .01$, $d = 0.83$.

Figure 1 displays the developmental course of the three types of mothers' contingent responsiveness and the emergence of culture-specific contingency patterns in the 3rd month. Intraindividual stabilities were moderate for the total contingency score ($r_{\text{Münster}} = .36$, $p > .10$, $r_{\text{Nso}} = .54$, $p < .05$) and visual contingencies ($r_{\text{Münster}} = .45$, $p < .10$, $r_{\text{Nso}} = .51$, $p < .05$), and mixed for auditory ($r_{\text{Münster}} = .45$, $p < .10$, $r_{\text{Nso}} = -.07$, $p > .10$) and proximal ($r_{\text{Münster}} = .37$, $r_{\text{Nso}} = .16$, both $ps > .10$) contingencies.

Provision of Face-to-Face Context and the 2-Month Shift

The percentage of 10-s intervals in which mothers established face-to-face context was subjected to a 2 (cultural group; between-subjects) × 5 (age; within-subjects) ANOVA. The mothers from the

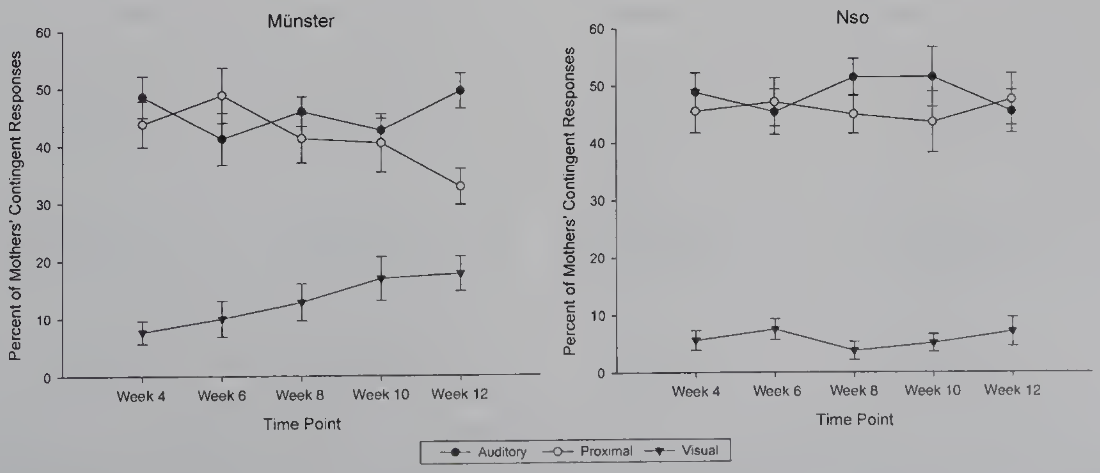


Figure 1. The developmental course of mothers' auditory, proximal, and visual contingencies for Münster and Nso infants.

Münster sample established face-to-face context significantly more often ($M = 0.82$, $SE = 0.05$), than did Nso mothers ($M = 0.49$, $SE = 0.04$), $F(1, 42) = 27.89$, $p < .01$, $\eta_p^2 = .40$. As indicated by the significant interaction, this difference increased with age, $F(4, 168) = 6.01$, $p < .01$, $\eta_p^2 = .13$ (see Table 2). Post hoc t tests indicate that mothers from Münster established face-to-face context for a significantly higher proportion of 10-s intervals than did Nso mothers in all weeks, $t_{\text{week } 4}(42) = 2.87$, $p < .01$, $d = 0.87$; $t_{\text{week } 6}(42) = 2.51$, $p = .016$, $d = 0.76$; $t_{\text{week } 8}(42) = 4.20$, $p < .001$, $d = 1.32$; $t_{\text{week } 10}(42) = 4.24$, $p < .001$, $d = 1.28$; and $t_{\text{week } 12}(42) = 8.58$, $p < .001$, $d = 2.78$.

Looking at the two cultural groups separately, one-way ANOVAs yielded a significant main effect of age for the Münster sample, $F(4, 76) = 3.24$, $p < .05$, $\eta_p^2 = .15$, that could best be explained by a linear increase, $F(1, 19) = 7.81$, $p < .05$, $\eta_p^2 = .29$. The significant main effect of age for the Nso sample, $F(4, 92) = 3.72$, $p < .01$, $\eta_p^2 = .14$, on the other hand, could best be explained by an inverted U-shaped function, $F(1, 23) = 8.64$, $p < .01$, $\eta_p^2 = .27$.

With respect to infants' alertness, there was a significant main effect of age, $F(4, 168) = 17.62$, $p < .01$, $\eta_p^2 = .30$. As indicated by the pairwise comparisons of estimated marginal means, there was a sharp increase in alertness between Weeks 6 and 8 for both sociocultural contexts; the respective mean differences for pairwise comparisons with Bonferroni adjustment ranged from −.18 to −.34, all $ps < .01$ (see Table 2).

The infants' gazing at their mothers' faces also differed between the cultural contexts and with age. During the episodes when mothers established face-to-face context (mutual gaze$_{\text{fav}}$), infants from the Münster sample looked at their mother faces, thus establishing eye contact, twice as long ($M = 0.36$, $SE = 0.04$) as did infants from the Nso sample ($M = 0.18$, $SE = 0.03$), $F(1, 42) = 14.57$, $p < .01$, $\eta_p^2 = .26$. There was a significant Cultural Context × Age interaction, $F(4, 168) = 4.21$, $p < .01$, $\eta_p^2 = .09$. Inspection of means (see Table 2) indicates that the difference between the cultural samples increased with age. Post hoc t tests yielded significant differences between the two cultural groups from Week 8 onward, $t_{\text{week } 8}(42) = 5.62$, $p < .001$, $d = 1.70$; $t_{\text{week } 10}(42) = 2.55$, $p = .015$, $d = 0.77$; and $t_{\text{week } 12}(42) = 3.06$, $p < .01$, $d = 0.93$.

The analysis of the duration of mutual gaze relative to the duration of the total interaction time (*mutual gaze*) yielded significant main effects for cultural group (Münster: $M = 0.28$, $SE = 0.03$; Nso: $M = 0.08$, $SE = 0.03$), $F(1, 42) = 24.50$, $p < .01$, $\eta_p^2 = .37$, as well as age, $F(4, 168) = 4.72$, $p < .01$, $\eta_p^2 = .10$. Furthermore, there was a significant Cultural Context × Age interaction, $F(4, 168) = 5.88$, $p < .01$, $\eta_p^2 = .12$. Looking at the two cultural groups separately, one-way ANOVAs yielded a significant main effect of age only for the Münster sample, $F(4, 76) = 5.67$, $p < .01$, $\eta_p^2 = .23$, that could best be explained by a sharp increase in mutual gaze between Weeks 6 and 8. Respective mean differences for pairwise comparisons with Bonferroni adjustment ranged between −.18 and −.23, all

Table 2

Face-to-Face Context, Alertness, and Mutual Gaze (Proportions) Across Time and Cultural Context

	Week 4	Week 6	Week 8	Week 10	Week 12
Face-to-face					
Münster	0.75 (0.26)	0.78 (0.27)	0.81 (0.20)	0.84 (0.23)	0.92 (0.10)
Nso	0.50 (0.31)	0.56 (0.30)	0.47 (0.32)	0.53 (0.25)	0.37 (0.27)
Alertness					
Münster	0.70 (0.30)	0.62 (0.35)	0.92 (0.16)	0.96 (0.12)	0.98 (0.04)
Nso	0.79 (0.35)	0.71 (0.41)	0.94 (0.20)	0.96 (0.10)	0.99 (0.02)
Mutual gaze$_{\text{fav}}$					
Münster	0.28 (0.26)	0.29 (0.25)	0.48 (0.26)	0.40 (0.33)	0.38 (0.25)
Nso	0.22 (0.22)	0.23 (0.29)	0.11 (0.15)	0.18 (0.23)	0.16 (0.22)
Mutual gaze					
Münster	0.16 (0.19)	0.16 (0.19)	0.39 (0.28)	0.36 (0.31)	0.35 (0.24)
Nso	0.09 (0.13)	0.09 (0.16)	0.06 (0.09)	0.10 (0.19)	0.08 (0.16)

Note. Values are M (SD). Scores are relative frequencies of 10-s intervals (face-to-face and alertness) or relative durations (mutual gaze and mutual gaze$_{\text{fav}}$). Mutual gaze$_{\text{fav}}$ indicates the duration of mutual gaze relative to the duration of episodes in which mothers established face-to-face context and infants were alert. Mutual gaze indicates the duration of mutual gaze relative to the duration of the entire episode. $N = 20$ for the Münster sample and $N = 24$ for the Nso sample.

ps $< .05$ (see Table 2). Post hoc *t* tests indicate that, from Week 8 on, dyads from the Münster sample established mutual gaze longer than did dyads from the Nso sample, $t_{\text{week }8}(42) = 5.18, p < .001, d = 1.57$; $t_{\text{week }10}(42) = 3.22, p < .01, d = 0.97$; and $t_{\text{week }12}(42) = 4.16, p < .001, d = 1.26$.

Intraindividual stabilities were reasonably high for the face-to-face context, $r_{\text{Münster}} = .49$, $r_{\text{Nso}} = .64$, both *ps* $< .05$, rather low for alertness, $r_{\text{Münster}} = .35$, $r_{\text{Nso}} = .17$, both *ps* $> .10$, and moderate for mutual gaze, $r_{\text{Münster}} = .33, p > .10$, $r_{\text{Nso}} = .45, p < .10$.

Contingency Patterns, Face-to-Face Context, and the 2-Month Shift

So far, the results indicate that changes in visual contingency parallel changes in the extent of face-to-face context provided by caretakers as well as changes in mutual gaze scores. For all of these scores, there was a significant increase across age for the Münster but not for the Nso sample. Within both cultural groups, averaged correlations between the two mutual gaze scores and visual contingency were significant, mutual gaze: $r = .50$, $p < .05$ for Münster and $r = .71$, $p < .01$ for Nso mothers, and mutual gaze$_{\text{fav}}$: $r = .41$, $p_{\text{one-sided}} < .10$ for Münster and $r = .64$, $p < .01$ for Nso mothers, supporting the assumption that on the dyad level, a longer duration of mutual gaze and mutual gaze$_{\text{fav}}$ was associated with a higher proportion of visual contingencies. Since, first, the two mutual gaze scores were highly correlated, $r = .95$ for the Münster sample and $r = .91$ for the Nso sample, *ps* $< .001$, and, second, there were less missing values for mutual gaze, we focused on mutual gaze in the following analyses.

For both mutual gaze and visual contingency, there were large differences between the two cultural groups from Week 8 onward. These data raise the question of whether the between-culture differences in mutual gaze and visual contingency were independent of each other or whether differences in visual contingency could, at least in part, be explained by differences in mutual gaze. The latter alternative would mean that if mothers and infants were to look at each other, mothers in both sociocultural contexts would behave similarly in terms of visual contingency. To address this issue, we computed two-step analyses separately for Weeks 8, 10, and 12 (van de Vijver & Leung, 1997). First, we computed a one-factorial (culture) ANOVA on visual contingency and, second, we computed a one-factorial (culture) ANCOVA on visual contingency with mutual gaze as the covariate. The change in η_p^2 for culture between the ANOVA and the ANCOVA gives us an indication of the extent to which differences in mutual gaze explain the cross-cultural differences in visual contingency. In all three ANOVAs, η_p^2s indicated large between-culture differences in visual contingency, smallest $F(1, 42) = 7.19$, $p < .05$, $\eta_p^2 > .14$. Entering mutual gaze as a covariate, the main effects of culture were no longer significant for any of the weeks, largest $F(1, 41) = .76$, $p > .10$, $\eta_p^2 < .04$. These findings support the hypothesis that mutual gaze largely explains cross-cultural differences in visual contingency.

These results indicate that during mutual gaze-episodes Nso mothers were as visually contingent as their Münster counterparts. The questions remain, however, as to why the Nso dyads generally engaged less in face-to-face interaction than did the Münster dyads and why the 2-month shift followed a different pattern (increase in awake alertness but not mutual gaze) in the Nso sample compared to the Münster sample. In order to address these questions, we conducted exploratory analyses of the intercorrelations between face-to-face context, mutual gaze, visual contingencies, and sociodemographic variables (e.g., mothers' education and age) that might help to explain intracultural variation. Of the sociodemographic variables that we analyzed, maternal education was the most relevant. We used a dichotomous score that differentiated between mothers who had received 7 years of formal schooling ($N = 17$) and mothers who had spent more time in school (either 9 [$N = 1$] or 12 [$N = 4$] years). Even though only five mothers had received a higher degree of formal education, averaged point-biserial correlations between education and all other scores were consistently positive, though not significant: $r = .25$ with face-to-face context, $r = .20$ with mutual gaze, and $r = .31$ with visual contingency. The correlations were highest between education and face-to-face context at 12 weeks, $r = .48$, $p < .05$; between education and mutual gaze at 4 weeks, $r = .50$, $p < .05$; and between education and visual contingency, $r_{\text{week }4} = .42$, $p < .10$; $r_{\text{week }10} = .47$, $p < .05$; $r_{\text{week }12} = .42$, $p < .10$. There were no systematic relations between education and either proximal, $r = -.26$ to $-.08$, $p > .10$ or auditory, $r = -.10$ to $.05$, $p > .10$, contingency scores.

Discussion

The purpose of this study was to analyze in detail how culture-specific contingency patterns develop during the first months of life. To address this issue, we explored how mothers' contingency patterns changed during the infants' 2nd and 3rd months of life and analyzed to what degree this development was influenced by sociocultural or other contextual factors.

As we hypothesized, there were no significant differences in the overall contingency level between cultural groups or with the infants' age. Mothers responded contingently to about 60% of their infants' NDVs. This result is very similar to Kärtner et al.'s (2008) finding of an overall contingency rate of 63% across six different sociocultural contexts. It seems, therefore, that contingent responsiveness is a universal component of intuitive parenting.

Also consistent with previous research is the finding that after 2–3 months of age, there were similarities as well as differences in the contingency patterns between the two cultural groups. Generally, mothers from both sociocultural contexts gave auditory responses far more often than they gave visual responses. Furthermore, in both samples there was no change in auditory contingencies with age. Other forms of maternal responsiveness, however, did vary across cultural contexts. As indicated by the proportional scores for auditory, proximal, and visual contingencies, 8- to 12-week-olds from the Münster sample experienced significantly more visual contingencies than did Nso infants and 12-week-olds from the Münster sample experienced significantly less proximal contingencies than did Nso infants. Within cultural contexts, 12-week-olds from the Münster sample experienced auditory contingencies more often than they did proximal contingencies, followed by visual contingencies. Twelve-week-olds from the Nso sample, in contrast, experienced proximal contingencies as often as they did auditory contingencies.

Interestingly, the modal differences in contingency patterns were not present at 4 weeks but emerged during the infants' 2nd and 3rd months. For the Münster sample, there was a linear increase in visual contingencies and a linear decrease in proximal contingencies. For the Nso sample, all contingency scores were continuous across the infants' age. Thus, these results clearly indicate differential developmental trajectories within the two sociocultural contexts.

In order to better understand this differential development, we analyzed the degree to which developmental processes were influenced by other factors, such as the provision of face-to-face context by caretakers and the infants' awake alertness and gazing pattern as indicators of the 2-month shift. As expected, mothers from the Münster sample established face-to-face context longer than did Nso mothers and this difference between the two cultural contexts increased with age. With respect to the 2-month shift, the infants' alertness increased sharply from Weeks 6 to 8 in both sociocultural contexts and reached about 100% in the 12th week.

The infants' gazing patterns are another indicator of the 2-month shift. In this respect, infants from the Münster and Nso samples were quite different. For the Münster infants, there was an age-dependent increase in the percentage of time that the infants looked back at their mothers while their mothers established face-to-face context (mutual gaze_{fav}). For the Nso infants, however, there was no change in the percentage of time that the infants looked back at their mothers. The difference in looking patterns with age between the two samples gave rise to significant differences between the samples from Weeks 8 to 12. As a consequence, the duration of mutual gaze relative to the total interaction time, a factor influenced by both the provision of face-to-face context and the infants' interest, exhibits an abrupt increase between Weeks 6 and 8 in the Münster sample but not in the Nso sample. This sharp increase in the duration of mutual gaze together with the sharp increase in alertness in the Münster sample can be taken as evidence of the 2-month shift as it is usually described in the literature (Lavelli & Fogel, 2002, 2005).

The results paint a different picture for the Nso infants since the 2-month shift seems to take a different shape: The Nso infants exhibited a similar sharp increase in awake alertness with age to the Münster infants, but the Nso infants' interest in their mothers' faces, that is, their gazing behavior, did not change with age. Across the 12 weeks, the relative duration of both face-to-face context established by the mothers and the two mutual gaze scores were continuous. Furthermore, interdyad differences remained relatively stable in the Nso context, and intercorrelations between face-to-face context, mutual gaze, and visual contingency were quite high for the Nso sample. These findings support the interpretation that each Nso mother provided a certain extent of face-to-face context and visual contingency that did not change much as the infant grew older.

In interpreting this cross-cultural developmental divergence, one has to keep in mind that most of the studies on the 2-month shift are monocultural studies from predominantly Western contexts (Emde, 1984; Lavelli & Fogel, 2002; Stern, 1985; Wolff, 1987). As indicated by the results, there was no difference in visual contingent responsiveness between the mothers from the two sociocultural contexts, per se. What differed between the two cultural contexts were the age-related changes in the provision of face-to-face context as well as the duration of mutual gaze itself. What are the possible reasons for these developmental differences in face-to-face context established by mothers and the infants' gazing behavior?

From a sociocultural perspective, we argue that caretakers hold culture-specific beliefs about parenting and follow different parenting strategies and behavioral scripts while interacting with their infants. These parenting beliefs and strategies lead to different developmental outcomes, which have an adaptive value within the specific sociocultural context. With respect to face-to-face context, there are several studies that have shown that face-to-face context and exclusive dyadic interaction play a central role in prototypically independent sociocultural contexts (Brazelton, 1977; Keller, 2007; Whiting & Pope Edwards, 1988). Furthermore, there is empirical evidence that distal parenting characterized by face-to-face interaction and object stimulation fosters the development of an autonomous self, which is one of the central developmental outcomes emphasized in independent sociocultural contexts (Keller, Kärtner, et al., 2005; Keller, Yovsi, et al., 2004). Nso mothers' parenting, in contrast, can best be described as proximal parenting with its emphasis on body contact and body stimulation (cf. Keller, 2007). According to Nso ethnotheories, these parenting strategies instill a sense of community that is characterized by obedience, deference, collective responsibility, and sharing (Goheen, 1996; Nsamenang & Lamb, 1994). Furthermore, there is empirical evidence that body contact and interactional warmth support the development of acceptance of the families' norms and values and compliance (Keller, Yovsi, et al., 2004; MacDonald, 1992).

It is important to keep in mind that, generally, infant communicative signals and mothers' behavior cannot be interpreted in isolation. Mother–infant interaction is a bidirectional process in which each partner's behavior is reciprocally influenced by the behavior of the other. Thus, the differences in the duration of gazing at the mothers' faces (mutual gaze$_{fav}$) found in this study cannot be attributed to the infants alone, even though we have described the infants' gazing as their *interest* in their mothers' faces. Rather, culture-specific scripts and behavioral standards of the mothers may have interacted with emerging behavioral potentials. Thus, in sociocultural contexts that favor face-to-face communication, mothers may enthusiastically promote mutual gaze and face-to-face interaction with their infants. Lavelli and Fogel (2005) identified positive feedback and mutual amplification as mechanisms that contribute to the consolidation of emerging behavioral patterns. In light of their results, the finding that the duration of time that the Münster infants looked at their mothers' faces while their mothers provided face-to-face context (mutual gaze$_{fav}$) increased parallel with an increase in the mothers' provision of face-to-face context may exemplify the consolidation of such an emerging behavioral pattern. For example, the more the mother establishes face-to-face context, the more probable it is that the infant will engage in mutual gaze if he or she is looking at his or her mother, which may be rewarding because of the stimulation that mothers provide during face-to-face interaction. For mothers, in turn, these experiences are rewarding because of their ethnotheoretical underpinning that mutual gaze and face-to-face interaction is a desirable way of interacting with infants. If this ethnotheoretical underpinning is missing or different, as in the Nso sample, stable and continuous development should be the consequence.

But what implications might the different ethnotheoretical underpinning of Nso mothers' caretaking have for the dynamics of mother–infant interaction? Generally, Nso mothers treat their infants as novices who need to learn compliance and subordination from early on (Nsamenang, 1992). In a qualitative analysis of the Week 12 interactions from the present study, Demuth (2008) described the prototypical Nso mother–infant interaction as characterized by rhythmic-synchronous structuring and both normative-hierarchical and socially oriented discourse. In contrast to the turn-taking sequences typical of mother–infant interaction in Münster, the Nso interactions include a large amount of repetitions and nonverbal vocalizations, which may lead to a symbiotic relationship stressing the unity between mother and infant. Repetitive patterns in Nso mother–infant interaction typically comprise musical chorusing as well as repeated addressing or greeting of the infant. These utterances were often synchronized with rhythmic body stimulation, for example, mothers rhythmically bouncing their infants. If one thinks in terms of consolidating behavioral patterns that are typical for interdependent sociocultural contexts, one may need to look beyond face-to-face interaction and toward developmental domains like affect regulation and calmness as an early precursor of proper demeanor and obedience (LeVine, 1994). However, the results of these studies point to the direction that, generally, Nso mothers take the lead and structure the interaction; that is, Nso parenting is more directive and is mother directed rather than child focused.

In the final step of our analyses, we explored to what extent other contextual factors (e.g., education) influenced the contingency patterns and dynamics of face-to-face interaction within the Nso sample. What we found was that the mothers' level of formal education (7 vs. 9 or more years) was associated with the provision of face-to-face context, the extent of mutual gaze, and the level of visual contingency. In a study on the effect of school attainment on mothers' contingent responsiveness, Richman et al. (1992) showed comparable results to ours. The higher the Mexican mothers' level of formal education (between 1 and 9 years), the more they talked and looked in response to vocalizations of their 10- or 15-month-olds. On the basis of these findings alone, it is unclear exactly how schooling affects maternal responsiveness. It seems, however, that schooling may provide women with experiences and models of social interaction that they would not acquire without schooling. These experiences may, in turn, influence the way in which they later interact with their infants. This interpretation is further supported by studies that have shown that education, with the accompanying changes in income and living standards, reorients socialization goals and parenting ethnotheories toward autonomy and independence (Laosa, 1982; LeVine, Miller, Richman, & LeVine, 1996). With regard to the Nso culture, Yovsi and Keller (2007) have shown that maternal school attainment is associated with changes in sleeping arrangements (shift from mother–infant to husband–wife cosleeping) and planned weaning date. Thus, there is evidence that formal education drives cultural change through its impact on ethnotheories, socialization goals, and norms of social interaction.

While the present study has a number of strengths in that it provided new evidence for intercultural similarities and differences in contingent responsiveness and the 2-month shift, it also has some limitations. First, due to the time-intensive coding of repeated observations, the sample sizes were relatively small. Furthermore, we substituted missing values in order to avoid listwise deletion in the repeated measures analyses. The robustness of the present findings, however, was supported by the fact that the pattern of significant results was identical regardless of whether missing values were substituted by mean scores, by adjacent values, or deleted listwise. Second, observations came from contexts that are prototypical for an independent and an interdependent sociocultural orientation: highly educated middle-class families in Münster and rural Nso farmers with only basic formal education living in a subsistence-based ecology. Based on our sociocultural approach, we argue that one has to be cautious about generalizing these data beyond the specific sociocultural contexts described. The contingency patterns that we found certainly do not apply to the country level. In order to determine whether generalizing these data to similar sociocultural contexts in the same or in other countries is appropriate, the results found here need to be replicated by further research with other prototypically independent or interdependent contexts. Last, there is evidence that Nso mothers focus more on proximal parenting; in Nso mothers' ethnotheories, mother–infant attachment is mainly defined in terms of bodily closeness, and body contact plays an important role in sensitive parenting (Otto, 2008; Yovsi, Kärtner, Keller, & Lohaus, 2009). By focusing on NDVs we tried to give justice to culture-specific emphases on face-to-face context and body contact. It is possible, however, that there are alternative mechanisms or communication systems that drive developmental change in early mother-infant communication in a way that is more typical for the Nso context.

In the present study, we could demonstrate that culture-specific contingency patterns in mother–infant interaction emerge during the 2nd and 3rd months of life. This differential development is associated with culture-specific differences in the duration of mutual gaze during social interaction, which indicates the effectiveness of broader cultural models, in this case, the prominence of face-to-face exchange in social interaction.

References

Bigelow, A. (1998). Infants' sensitivity to familiar imperfect contingencies in social interaction. *Infant Behavior & Development*, *21*, 149–161.

Bigelow, A. E., & DeCoste, C. (2003). Sensitivity to social contingency from mothers and strangers in 2-, 4-, and 6-month-old infants. *Infancy*, *4*, 111–140.

Bigelow, A. E., & Rochat, P. (2006). Two-month-old infants' sensitivity to social contingency in mother–infant and stranger–infant interaction. *Infancy*, *9*, 313–325.

Bloom, K., Russell, A., & Wassenberg, K. (1987). Turn taking affects the quality of infant vocalizations. *Journal of Child Language, 14*, 211–227.

Bornstein, M., Tamis-LeMonda, C., Tal, J., Ludemann, P., Toda, S., Rahn, M., et al. (1992). Maternal responsiveness to infants in three societies: The United States, France, and Japan. *Child Development, 63*, 808–821.

Brazelton, T. B. (1977). Implications of infant development among the Mayan Indians of Mexico. In P. H. Leiderman, S. R. Tulkin, & A. Rosenfeld (Eds.), *Culture and infancy: Variations in the human experience* (pp. 151–188). New York: Academic Press.

Central Intelligence Agency. (2008). *The world factbook 2002*. Retrieved September 30, 2008, from http://www.umsl.edu/services/govdocs/wofact2004/

Demuth, D. (2008). *Talking to infants: How culture is instantiated in early mother-infant interactions. The case of Cameroonian farming Nso and North German middle-class families*. Unpublished doctoral dissertation, University of Osnabrück, Germany.

Emde, R. N. (1984). The affective self: Continuities and transformations from infancy. In J. D. Call, E. Galenson, & R. L. Tyson (Eds.), *Frontiers of infant psychiatry* (pp. 38–54). New York: Basic Books.

Fogel, A., Toda, S., & Kawai, M. (1988). Mother-infant face-to-face interaction in Japan and the United States: A laboratory comparison using 3-month-old infants. *Developmental Psychology, 24*, 398–406.

Goheen, M. (1996). *Men own the fields, women own the crops: Gender and power in the Cameroon grassfields*. Madison: University of Wisconsin Press.

Haith, M. M., Bergman, T., & Moore, M. J. (1977). Eye contact and face scanning in early infancy. *Science, 198*(4319), 853–855.

Hsu, H.-C., & Fogel, A. (2003). Social regulatory effects of infant nondistress vocalization on maternal behavior. *Developmental Psychology, 39*, 976–991.

Hsu, H.-C., Fogel, A., & Messinger, D. S. (2001). Infant non-distress vocalization during mother-infant face-to-face interaction: Factors associated with quantitative and qualitative differences. *Infant Behavior and Development, 24*, 107–128.

Jaffe, J., Beebe, B., Stanley, F., Crown, C., & Jasnow, M. (2001). Rhythms of dialogue in infancy: Coordinated timing in development. *Monographs of the Society for Research in Child Development, 66*(2, Serial No. 131).

Kärtner, J., Keller, H., Lamm, B., Abels, M., Yovsi, R., Chaudhary, R., et al. (2008). Similarities and differences in contingency experiences of 3-month-olds across sociocultural contexts. *Infant Behavior and Development, 31*, 488–500.

Keller, H. (2006). Germany: Continuity and change. In J. Georgas, J. Berry, F. Van de Vijver, C. Kağitçibaşi, & Y. H. Poortinga (Eds.), *Families across cultures* (pp. 327–335). Cambridge, UK: Cambridge University Press.

Keller, H. (2007). *Cultures of infancy*. Mahwah, NJ: Erlbaum.

Keller, H., Hentschel, E., Yovsi, R., Lamm, B., Abels, M., & Haas, V. (2004). The psycho-linguistic embodiment of parental ethnotheories: A new avenue to understanding cultural processes in parental reasoning. *Culture and Psychology, 10*, 293–330.

Keller, H., Kärtner, J., Borke, J., Yovsi, R., & Kleis, A. (2005). Parenting styles and the development of the categorical self: A longitudinal study on mirror self-recognition in Cameroonian Nso and German families. *International Journal of Behavioral Development, 29*, 496–504.

Keller, H., Lohaus, A., Völker, S., Cappenberg, M., & Chasiotis, A. (1999). Temporal contingency as an independent component of parenting behavior. *Child Development, 70*, 474–485.

Keller, H., Otto, H., Lamm, B., Yovsi, R., & Kärtner, J. (2008). The timing of verbal/vocal communications between mothers and their infants: A longitudinal cross-cultural comparison. *Infant Behavior and Development, 31*, 217–226.

Keller, H., & Schölmerich, A. (1987). Infant vocalizations and parental reactions during the first 4 months of life. *Developmental Psychology, 23*, 62–67.

Keller, H., Voelker, S., & Yovsi, R. (2005). Conceptions of parenting in different cultural communities: The case of West African Nso and Northern German women. *Social Development, 14*, 158–180.

Keller, H., Yovsi, R., Borke, J., Kärtner, J., Jensen, H., & Papaligoura, Z. (2004). Developmental consequences of early parenting experiences: Self-recognition and self-regulation in three cultural communities. *Child Development, 75*, 1745–1760.

Keller, H., Zach, U., & Abels, M. (2005). The German family: Families in Germany. In J. L. Roopnarine & U. P. Gielen (Eds.), *Families in global perspective* (pp. 242–258). Boston: Pearson.

Laosa, L. M. (1982). School, occupation, culture, and family: The impact of parental schooling on the parent-child relationship. *Journal of Educational Psychology, 74*, 791–827.

Lavelli, M., & Fogel, A. (2002). Developmental changes in mother-infant face-to-face communication: Birth to 3 months. *Developmental Psychology, 38*, 288–305.

Lavelli, M., & Fogel, A. (2005). Developmental changes in the relationship between the infant's attention and emotion during early face-to-face communication: The 2-month transition. *Developmental Psychology, 41*, 265–280.

LeVine, R. A. (1994). *Child care and culture: Lessons from Africa*. Cambridge, UK: Cambridge University Press.

LeVine, R. A., Miller, P. M., Richman, A. L., & LeVine, S. (1996). Education and mother-infant interaction. A Mexican case study. In S. Harkness & C. M. Super (Eds.), *Parents' cultural belief systems: Their origins, expressions, and consequences* (pp. 254–288). New York: Guilford.

LeVine, R. A., & Norman, K. (2001). The infant's acquisition of culture: Early attachment re-examined in anthropological perspective. In C. C. Moore & H. F. Mathews (Eds.), *The psychology of cultural experience* (pp. 83–104). Cambridge, UK: Cambridge University Press.

MacDonald, K. B. (1992). Warmth as a developmental construct: An evolutionary analysis. *Child Development, 63*, 753–773.

Nsamenang, A. B. (1992). *Human development in cultural context: A third world perspective.* Newbury Park, CA: Sage.

Nsamenang, A. B., & Lamb, M. (1994). Socialization of Nso children in the Bamenda Grassfields of Northwest Cameroon. In P. Greenfield & R. Cocking (Eds.), *Cross-cultural roots of minority child development* (pp. 133–146). Hillsdale, NJ: Erlbaum.

Ogunnaike, O., & Houser, R. (2002). Yoruba toddlers' engagement in errands and cognitive performance on the Yoruba Mental Subscale. *International Journal of Behavioral Development, 26*, 145–153.

Otto, H. (2008). *Culture-specific attachment strategies in the Cameroonian Nso: Cultural solutions to a universal developmental task.* Unpublished doctoral dissertation, University of Osnabrück, Germany.

Papoušek, H., & Papoušek, M. (1987). Intuitive parenting: A dialectic counterpart to the infant's integrative competence. In J. D. Osofsky (Ed.), *Handbook of infant development* (2nd ed., pp. 669–720). New York: Wiley.

Richman, A., Miller, P., & LeVine, R. (1992). Cultural and educational variations in maternal responsiveness. *Developmental Psychology, 28*, 614–621.

Rochat, P. (2001). *The infant's world.* Cambridge, MA: Harvard University Press.

Statistisches Bundesamt. (2008). *Durchschnittliches Alter der Mütter bei der Geburt und Durchschnittliche Kinderzahl je Frau.* Retrieved September 26, 2008, from http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Navigation/Statistiken/Bevoelkerung/GeburtenSterbefaelle/GeburtenSterbefaelle.psml

Stern, D. (1985). *The interpersonal world of the infant.* New York: Basic Books.

Symons, D., & Moran, G. (1994). Responsiveness and dependency are different aspects of social contingencies: An example from mother and infant smiles. *Infant Behavior and Development, 17*, 209–214.

Tarabulsy, G., Tessier, R., & Kappas, A. (1996). Contingency detection and the contingent organization of behavior in interactions: Implications for socioemotional development in infancy. *Psychological Bulletin, 120*, 25–41.

van der Kloot, W. (1998). Univariate and multivariate analysis of variance of longitudinal data. In C. C. J. H. Bijleveld & L. J. Van der Kamp (Eds.), *Longitudinal data analysis: Designs, models, & methods* (pp. 155–206). Thousand Oaks, CA: Sage.

van de Vijver, F., & Leung, K. (1997). *Methods and data analysis of cross-cultural research.* Thousand Oaks, CA: Sage.

Van Egeren, L., Barratt, M., & Roach, M. (2001). Mother-infant responsiveness: Timing, mutual regulation, and interactional context. *Developmental Psychology, 37*, 684–697.

Watson, J. S. (1985). Contingency perception in early social development. In T. M. Field & N. A. Fox (Eds.), *Social perception of infants* (pp. 157–176), Norwood, NJ: Ablex.

Whiting, B. B., & Pope Edwards, C. (1988). *Children of different worlds: The formation of social behavior.* Cambridge, MA: Harvard University Press.

Wolff, P. H. (1987). *Behavioral states and the expression of emotion in early infancy.* Chicago: University of Chicago Press.

Yovsi, D. R. (2003). *Ethnotheories on breastfeeding and mother-infant interaction. The case of the Nso and Fulani mothers and their infants 3–6 months of age, in Mbven subdivision of the North-West Province of Cameroon.* Münster, Germany: LIT Verlag.

Yovsi, D. R., Kärtner, J., Keller, H., & Lohaus, A. (2009). Maternal interactional quality in two cultural environments: German middle class and Cameroonian Nso mothers. *Journal of Cross-Cultural Psychology, 40*, 701–707.

Yovsi, D. R., & Keller, H. (2007). The architecture of co-sleeping among wage-earning and subsistence farming Cameroonian Nso families. *Ethos, 35*, 65–84.

Child Development, March/April 2010, Volume 81, Number 2, Pages 555–567

"My Way or Mom's Way?" The Bilingual and Bicultural Self in Hong Kong Chinese Children and Adolescents

Qi Wang and Yi Shao
Cornell University

Yexin Jessica Li
Arizona State University

This study examined the relation of language to the development of a cultural self. Bilingual children ages 8–14 from Hong Kong ($N = 125$) were interviewed in either English or Chinese. They recalled autobiographical events and described themselves, and indicated their agreement with Chinese interdependent versus Western independent values. Children interviewed in English provided more elaborate and self-focused self-descriptions and memory accounts and endorsed more strongly Western values, compared with children interviewed in Chinese. Furthermore, the endorsement of a cultural belief system mediated the effect of language on self-concept, which, in turn, mediated the effect of language on autobiographical memory. These findings offer new insight into the dynamic relations between language, culture, and the self.

The self is a multifaceted, complex, dynamic construct encompassing many interacting aspects or components (Baumeister, 1998; Neisser, 1988). Two aspects of the self—autobiographical memory and self-concept—have been underscored in contemporary theories of the self (Fivush & Nelson, 2004; Harter, 1999; Nelson, 2001). Autobiographical memory, or the "extended self," refers to memory for significant personal experiences from an individual's life. Self-concept, or the "conceptual self," refers to an individual's conceptual representations of him- or herself (Neisser, 1988). While autobiographical memory provides the materials for the constitution and maintenance of a dynamic self-concept, self-concept exerts executive control over the constructive process of autobiographical remembering in accordance with the self's current goals and motivations and thus acts to "shape both the accessibility of memories and the accessibility of their content" (Conway, 2005, p. 595; Conway & Pleydell-Pearce, 2000; Ross, 1989). More important from the current perspective, the two aspects of the self are both sensitive to social-cultural influences as they develop in a myriad of daily interactions between individuals and their immediate and distal contexts, together making up a self that integrates the framework of the culture, namely, the cultural self (see Wang, 2006a, for a review).

The Cultural Self

The development of the self is immersed in the social-cultural context constituted by cultural beliefs, symbols, artifacts, metaphors, and everyday practices (Bruner, 1990; Miller, Fung, & Koven, 2007; Wang, 2006a). These cultural factors provide an important framework based on which individuals process and represent self-related information (Kağitçibaşi, 2007; Markus & Kitayama, 1991). The belief system in European American culture that emphasizes individuality, personal uniqueness, and self-sufficiency promotes an autonomous self-concept that is primarily defined by one's inner attributes, qualities, and opinions. This autonomous self-concept may further direct cognitive resources toward elaborate processing of event information that highlights one's uniqueness and autonomy, and such information is likely to become richly represented and highly accessible during recall. In contrast, the emphasis on interdependence, group solidarity, and relational hierarchy in Confucian cultures such as Japan, Korea, and China encourages the development of a relational self-concept that is largely defined by one's position in a nexus of social relations. This relational self-concept may motivate individuals to focus on event information about group activities and interactions, which helps

---

This research was supported by grants from the Cornell East Asia Program and the U.S. Department of Agriculture to the first author. We thank Charis Wing Man Law, Bernard Lam, and May Wong for their assistance. We also thank Françoise Vermeylen for statistical consultation. Special thanks go to the participants who made this study possible.

Correspondence concerning this article should be addressed to Qi Wang, Department of Human Development, Cornell University, MVR Hall, Ithaca, NY 14853-4401. Electronic mail may be sent to qw23@cornell.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8102-0012

them relate to significant others and the community (Wang, 2006a; Wang & Ross, 2007).

Consistent with these theoretical views, research has shown that when asked to describe themselves, European American individuals, both children and adults, tend to focus on their unique attributes, qualities, and beliefs independent of others or social contexts, whereas Japanese, Koreans, and Chinese frequently refer to their social roles, group memberships, and relationships with significant others (Cousins, 1989; Rhee, Uleman, Lee, & Roman, 1995; Trafimow, Triandis, & Goto, 1991; Wang, 2001, 2004). These different self-concepts across cultures are further related to different ways of autobiographical remembering, whereby European American children and adults often provide elaborate, lengthy accounts of personal experiences and focus on their own roles, feelings, and perspectives, whereas the memories of Koreans and Chinese are often brief and centered on shared activities and significant others (Han, Leichtman, & Wang, 1998; Mullen, 1994; Wang, 2001, 2004; Wang & Leichtman, 2000). These cross-cultural findings support the notion that the development of the self is a process of cultural adaptation (Kağitçibaşi, 2007; Miller et al., 2007; Wang, 2006a), during which individuals come to internalize cultural belief systems (e.g., independence vs. interdependence) into their own self-understanding and remembering (e.g., with an individual vs. a social orientation). Importantly, the acquisition of the cultural self may be attributed to a host of sociocultural variables, among which language as a symbolic system of culture may be central.

Language and the Self

Language lends stability to human experience and at the same time sustains flexibility adaptive to specific contexts (Nelson, 2001). Indeed, many theorists argue that the self can be viewed as developed, expressed, and reconstructed through narrative practices that occur in the realm of linguistic formulation (Bruner, 1990; Fivush & Nelson, 2004; Ochs, 1996). In particular, linguistic representation, through the structure and lexicon of the language and the social arrangement of the communicative context, may direct individuals' attention to particular cultural models of self and render certain qualities or representations of the self salient and accessible (Chiu, Leung, & Kwan, 2007; Holland & Quinn, 1987). Furthermore, as proposed within the language socialization approach, language plays an essential role in the communication of cultural content while, in the meantime, contributing to meaning construction through linguistic forms (Fernald & Morikawa, 1993; Miller et al., 2007; Ochs, 1996). Language learning is thus a powerful medium of cultural transmission that enables children to acquire both linguistic and sociocultural competence.

One important approach to understanding the relation between culture, language, and the self is to study bilingual individuals. Given that cultural practices and meanings are so deeply embedded in language, the acquisition of a second language often entails the establishment of a new cognitive system—schemata, mental models, or representations—that reflects new ways of constructing the self and its relation to the social and physical world (Ochs, 1996; Schrauf, 2000; Valsiner, 2001). As a result, bilingual individuals may possess separate self-structures associated with different languages. Studies with bilingual adults have provided some evidence. Ross, Xun, and Wilson (2002) interviewed, in either Chinese or English, Chinese-born bilingual young adults living in Canada. They asked participants to describe themselves and to rate statements reflecting Chinese cultural values (e.g., "Modesty leads to success, pride leads to failure"). Those who participated in Chinese described themselves in more collective terms and endorsed Chinese values more strongly than did those surveyed in English. Similarly, when bilingual young adults from Hong Kong were asked to describe themselves, those who responded in Chinese used more collective and fewer individual terms than those who responded in English (Kemmelmeier & Cheng, 2004; Trafimow, Silverman, Fan, & Law, 1997). In addition, in a study of autobiographical memory in Russian English bilingual adults, Marian and Kaushanskaya (2004) found that memories recalled in English tended to be more self-focused, whereas memories recalled in Russian were more other oriented. These effects of language on self-concept and memory directly parallel the findings of cross-cultural studies (Han et al., 1998; Mullen, 1994; Wang, 2001, 2004; Wang & Ross, 2005).

From a developmental perspective, it is of great importance to examine whether the acquisition of different languages in children is accompanied by different modes of self-understanding and remembering. Furthermore, although the findings with adults suggest that language can bring self-concept and autobiographical memory in line with cultural belief systems, the specific mechanisms underlying the language effects are yet to be fully explicated. According to the constructivist view, language may serve as a cue or prime for associated cultural constructs and thus increase the accessibility of relevant beliefs, self-concepts, and autobiographical information (Ross et al., 2002; Schrauf, 2000; Trafimow et al., 1997). More important from the current perspective, these language effects may not reflect simultaneous processes. Conceivably, the use of a particular language (e.g., Chinese) may activate a specific cultural frame or belief system (e.g., Confucian values) to which this language is chronically connected in everyday life. The cultural belief system may, in turn, activate culturally promoted self-concepts (e.g., a relational self), which may further increase the accessibility of relevant autobiographical event information (e.g., social interactions). In other words, the activated cultural belief system may mediate the effect of language on self-concept. In turn, given its executive role of modulating access to long-term memory (Conway, 2005; Conway & Pleydell-Pearce, 2000), the working self-concept may further mediate the effect of language on autobiographical memory. Figure 1 depicts the hypothesized direct effects and mediations. Alternatively, the cultural accommodation theory suggests that language may create a "response bias," where individuals provide responses deemed appropriate or desirable by the culture in which the language is used (Bond & Yang, 1982). Thus, the use of a particular language may elicit associated culturally desirable beliefs, self-concept, and autobiographical recall. Such effects should appear in parallel and the unidirectional mediations will not be expected. This study addresses these theoretical issues by examining self-concept and autobiographical memory in bilingual children and adolescents from Hong Kong.

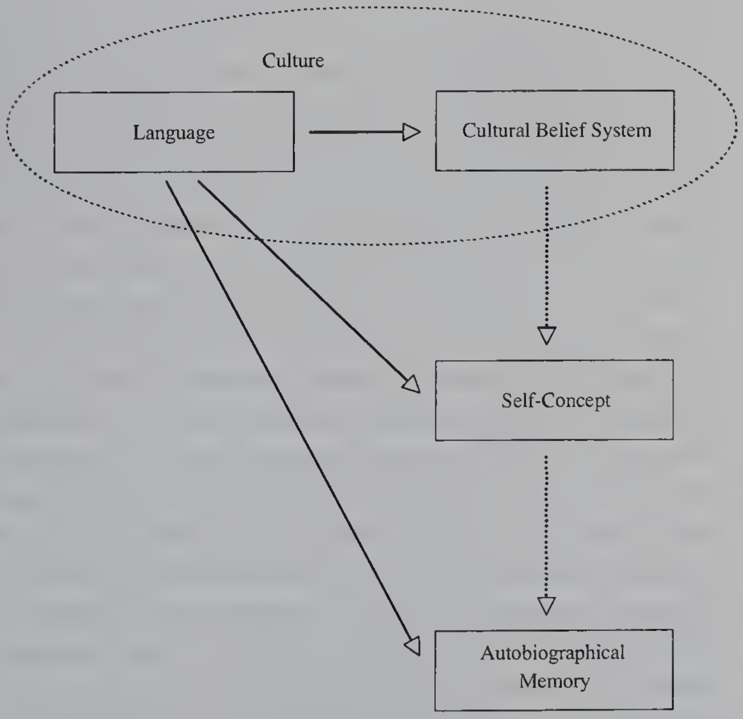


Figure 1. The effects of language.
Note. The solid arrows represent the direct effects of language predicted by both the constructivist view and the cultural accommodation theory. The dashed arrows represent hypothesized mediations according to the constructivist view.

The Language and Culture of Hong Kong Children

Because of its 100-year reign by England that lasted until 1997, Hong Kong is a culturally dynamic place where Western and Chinese values intertwine and where modernity and tradition coexist. Historically, the British colonial and metropolitan authorities in Hong Kong promoted English-language education (vis-à-vis Chinese education). The policies and practices embrace both the promotion of Western education through the medium of English and the teaching and learning of English as a language subject (Tung, Lam, & Tsang, 1997). After its handover back to China, Hong Kong maintains its cultural ties to the Western world under the Chinese government's "one country, two systems" policy (Hong, Morris, Chiu, & Benet-Martinez, 2000). Both Chinese and English are official languages. Many schools that were in place during the British reign have remained relatively unaffected, and English continues to be an important language to be learned starting even before formal schooling.

On the other hand, the Hong Kong society is conscious of preserving Chinese traditions, where Confucian ethics such as respect for the elderly, humility, and a sense of shame are highly prized. Children are exposed to traditional Chinese culture early on and come to learn and speak Chinese (Cantonese or Mandarin, depending on their family background) in their daily lives at home and in the community. Some parents also send their children to take additional Chinese classes outside school. Consequently, many children become proficient bilinguals from an early age. Furthermore, constant transition in everyday life between different cultural contexts such as home and school often requires full cultural competency in both (Hong et al., 2000; Kemmelmeier & Cheng, 2004). Many children thus grow up bicultural, living with their families in a city that has established traditional Chinese teachings and a rich cultural history, while seeking out Western friends, music, movies, toys and cuisine that are readily available to them (Deeds, Stewart, Bond, & Westrick, 1998; Feldman

& Rosenthal, 1991). The bilingual and bicultural characteristics of Hong Kong children thus present us with a unique opportunity to test the relation of language to the development of a cultural self.

The Present Study

In the present study, bilingual 8- to 14-year-old children and adolescents from Hong Kong were interviewed in either English or Chinese. They were asked to recall autobiographical events and provide self-descriptions, and to indicate their agreement with Chinese interdependent versus Western independent values. We expected that, consistent with previous research with adults (Marian & Kaushanskaya, 2004; Ross et al., 2002; Trafimow et al., 1997), children interviewed in English would report higher agreement with Western independent values and lower agreement with Chinese values than those interviewed in Chinese, regardless of age. Also, across all age groups, children interviewed in English were expected to provide lengthier and more self-focused descriptions of themselves and their personal experiences, compared with those interviewed in Chinese whose memories and self-descriptions would be less elaborate and more socially oriented. Furthermore, in line with the constructivist view, we expected that the language of interview would prime the associated cultural belief system, which would then activate the self-concept consistent with the belief system and further shape the retrieval of autographical event information. Thus, the endorsement of a cultural belief system would mediate the effect of language on self-concept, which, in turn, would serve as a mediator for the effect of language on autobiographical memory.

Furthermore, including different age groups would help to reveal possible interactions between age, language, and cultural identification. Extant cross-cultural studies on the development of self-concept and autobiographical memory have focused on preschool and early grade school children (Han et al., 1998; Wang, 2004), included only one age group (Stigler, Smith, & Mao, 1985; Wang, 2006b; Wang & Leichtman, 2000), or not examined age effect in analysis (e.g., Hart, Lucca-Irizarry, & Damon, 1986). In Wang (2004) that included 4-, 6-, and 8-year-olds from European American and Chinese cultural backgrounds, it was found that older children in both cultures provided more self-descriptions and lengthier memories than younger children, which reflects children's increasing cognitive skills with age. Accordingly, compared with younger children, older children described more personal as well as social themes in their memory recall and self-descriptions. Moreover, there were interactions between culture and age such that cultural differences, while identified in all age groups, appeared to become larger with age, suggesting the increasing enculturation as children grow older. It would be interesting to see whether similar patterns of results with the younger, monocultural children would be found in the older, bicultural children in the present study. However, given that there is no extant research on relevant topics with bilingual children in middle childhood and early adolescence, we made no priori predictions about effects related to age.

In addition, we examined gender effects because literature has suggested that across cultures, females often exhibit greater interdependence, whereas males often exhibit greater independence (Cross & Madson, 1997; Fivush & Buckner, 2003). Such gender differences might be reflected in children's value orientations and self-knowledge such that, independent of language of interview, girls would endorse more strongly interdependent as opposed to independent values and provide more socially oriented self-descriptions and memory accounts, when compared with boys.

Method

Participants

The participants were 125 children from Hong Kong, including thirty-three 8-year-olds ($M = 7.62$, $SD = 0.76$, $Mdn = 7.80$; 20 girls), thirty-two 10-year-olds ($M = 10.02$, $SD = 0.48$, $Mdn = 10.00$; 13 girls), twenty-eight 12-year-olds ($M = 12.02$, $SD = 0.58$, $Mdn = 12.03$; 14 girls), and thirty-two 14-year-olds ($M = 14.58$, $SD = 0.98$, $Mdn = 14.47$; 22 girls). All children were ethnic Chinese and were proficient speakers of both English and Chinese (Mandarin or Cantonese). Based on initial conversations with the children, research assistants rated each child's proficiency in both languages, respectively, in comparison with native speakers, on a 10-point scale of 1 (*very poor*) to 10 (*excellent*). The average Chinese proficient score was 9, and the average English proficient score was 8. There was no age difference in either score ($Fs < 1$, *ns*). Children were recruited from elementary and middle schools in primarily middle- and upper-middle-class communities. All parents gave informed consent. Fifty percent of the children's mothers had a college education or beyond, 38% had a high school education, 2% received postgraduate professional training, and

Table 1
Mean Ages (Standard Deviation; % Girls) by Interview Condition

Age group (years)	Chinese	English
8	7.57 (0.71; 61)	7.71 (0.90; 60)
10	9.97 (0.45; 42)	10.17 (0.56; 38)
12	11.93 (0.69; 47)	12.12 (0.42; 54)
14	14.51 (1.20; 71)	14.63 (0.81; 67)

Note. Within each age group, there was no significant difference in age or gender distribution between the two language conditions.

10% did not provide the information. Children were randomly assigned to the Chinese or English interview condition, being matched as closely as possible on the basis of age and gender. Table 1 displays the mean ages and gender distributions of the children by interview condition. The Chinese interviews were conducted in Mandarin (11%) or Cantonese (89%), whichever dialect the child was most comfortable with.

Procedure

Trained bilingual research assistants visited children at home and conducted the interview in a quiet room without the presence of parents or other family members. The interview procedure followed previous research on children's autobiographical memory and self-concept (Han et al., 1998; Wang, 2004). At the beginning of the interview, the interviewer first chatted with the child to establish rapport. When the child seemed relaxed and comfortable, the interviewer told the child, "You and I are going to play a fun game. It's called a question-and-answer game. I'm going to ask you some questions. You can reply in any way you want. There are no right or wrong answers. You'll see it's really fun. Are you ready?"

The first set of questions was about children's memories of personal experiences. Children were asked to recount four past events, including: one thing they did recently that was special and fun, a recent time when they argued with someone at school, a recent time when they did well on a school assignment, and their earliest childhood memory. After each memory question, the interviewer used standard prompts such as "What else happened?" and "Can you tell me more about it?" until the child indicated by speech or gesture that the memory was finished.

Following the memory questions, children were asked to describe themselves in response to open-ended questions (Keller, Ford, & Meacham, 1978; Wang, 2004). The interviewer told the child: "(Child's name), I would like to write a story about you, to write a story that will tell about (child's name). What's the first thing I should put in the story?" She prompted the child after each response, "And what else should I write to tell about you?" until the child indicated by speech or gesture that he or she was finished.

The last set of questions tapped upon children's identification with or endorsement of Chinese cultural values such as interpersonal harmony, group solidarity, social compliance, and humility, versus Western independent values such as individual autonomy, personal sufficiency, and pride. The questions were adapted from Ross et al. (2002), the Self-Construal Scale (Singelis, 1994), and the Ethnic Identity Scale (Umaña-Taylor, Yazedjian, & Bámaca-Gómez, 2004) and modified to be age appropriate. The interviewer posed to children 20 pairs of opposing statements that described a variety of behaviors, preferences, and emotional reactions pertaining to important others, social situations, and cultural artifacts. For example, children were asked, "If you get the highest grade in class, will you tell people about it or keep it to yourself?" and "When you have a big decision to make, do you ask your parents for advice or make your own plans?" Children responded by making a forced choice for each pair of the statements.

The order of the tasks was fixed across all children in order to test the hypothesized mediations in the most stringent way because any possible sequence effects (i.e., autobiographical recall → self-descriptions → culture beliefs) would work against the mediations (i.e., culture beliefs → self-descriptions → autobiographical recall). Each interviewer–child session lasted approximately 30 min. Children received small gifts for their participation. All interviews were audio tape-recorded and then transcribed verbatim onto paper for coding.

Coding

Endorsement of cultural values. Children's responses to the cultural value questions were scored 1 whenever they agreed with Chinese values (e.g., asking parents for advice) and scored 0 whenever they responded in accordance with Western independent values (e.g., making one's own plans). Reliability analysis showed that six items had negative correlations with the total scale (rs = −.04 to −.38) and they were therefore excluded. The remaining items formed a reliable scale (Cronbach's α = .70), and their scores were aggregated to index the children's endorsement of Chinese values (maximum score 14). Note that because the total number of items was fixed (i.e., 14 items), a low score on the scale (e.g., 3) would indicate low agreement with Chinese interdependent values (i.e., responding to 3 items in accordance with Chinese values) and high agreement with Western independent values (i.e., responding to 11 items in accordance with Western values). In other words, if each child were to receive two separate scores, one for the number of responses in accordance with Chinese values and one for the number of responses in accordance with Western values, the two scores would be perfectly correlated at −1 and they would yield identical results in analysis, only in different directions.

Self-concept. Children's self-descriptions were coded as referring to personal versus social aspects of the self, following previous research (Trafimow et al., 1991; Wang, 2001, 2004). Proposition, defined as a subject–verb construction (in Chinese, 主谓结构), was used as the coding unit (Fivush, Haden, & Adam, 1995). Each unique or implied verb in an independent clause forms a propositional unit (e.g., ''I normally go to the library''). The total number of self-descriptions each child provided was first tabulated. Then self-descriptions were coded as *personal* whenever children referred to personal qualities, attitudes, beliefs, and behaviors that were not related to other people (e.g., ''I am honest'' and ''I enjoy books''). Self-descriptions were coded as *social* whenever children referred to their social roles, demographic categories, and group memberships (e.g., ''I live in Hong Kong'' and ''I am a student.''), or important others, relationships, and sensitivity to the viewpoints of others (e.g., ''I have many friends,'' ''My sister is at high-school,'' and ''I like to help other people''). The two codes were exclusive and exhaustive. A ''personal-to-social self-description'' ratio was then calculated for each child to form a composite score of *autonomous self*, which controlled for the effect of the total number of self-descriptions. Some children did not provide any social self-descriptions, a constant 1 was therefore added to the denominator. The ratio scores were later submitted to analysis. Two trained research assistants, both blind to the hypotheses, first coded 20% of the data for reliability estimates, which ranged from 81% to 88%. Disagreement was resolved through discussion between the coders. One assistant then coded the rest of the data.

Autobiographical memory. Following previous studies (Han et al., 1998; Wang, 2004), the length of children's memories was first coded by counting the total number of English words or Chinese characters children used in each memory. Memory content was then coded, using proposition as the coding unit (Fivush et al., 1995; e.g., ''Yesterday I went to fetch my award''). Propositions were coded as *personal* whenever children referred to their own emotions, actions, thoughts, preferences, and agency in their memories (e.g., ''I was happy'' and ''I wanted to drink Coca Cola''). Propositions were coded as *social* whenever children referred to social interactions, group activities, and the roles of others (e.g., ''We traveled'' and ''My father bought a Game Cake for me''). The two codes were exclusive but not exhaustive. A ''personal-to-social'' ratio was then calculated for each memory to index the degree of *memory self-focus* while controlling for the effect of memory length. A constant 1 was added to the denominator because some children did not describe any social content in their memories. The ratio scores were later submitted to analysis. Two trained research assistants first coded 20% of the data for reliability estimates, which ranged from 82% to 96%. Disagreement was resolved through discussion, and one assistant then coded the rest of the data.

Results

In connection with the hypotheses, results pertaining to the endorsement of cultural values, self-concept, and autobiographical memory are first presented. Across all the variables, 2 (interview language: Chinese vs. English) × 2 (gender: boys vs. girls) × 4 (age: 8-year-olds vs. 10-year-olds vs. 12-year-olds vs. 14-year-olds) analyses of variance (ANOVAs) were conducted. Power analysis ($\alpha = .05$, power = 0.80, and effect size $f = 0.30$) indicated that the current sample size was adequate to test the three-factor model (Rosenthal & Rosnow, 1991). Children's responses to the four memory questions showed overall consistent patterns; the analyses of memory variables were therefore performed based on the mean frequencies across the four memories. Mediation analyses were then conducted to test the hypotheses that the language use would activate the associated cultural belief system, which would mediate the effect of language on self-concept. Self-concept, in turn, would serve as a mediator for the effect of language on autobiographical memory. Some children did not answer every question; thus, the degrees of freedom varied slightly across tests.

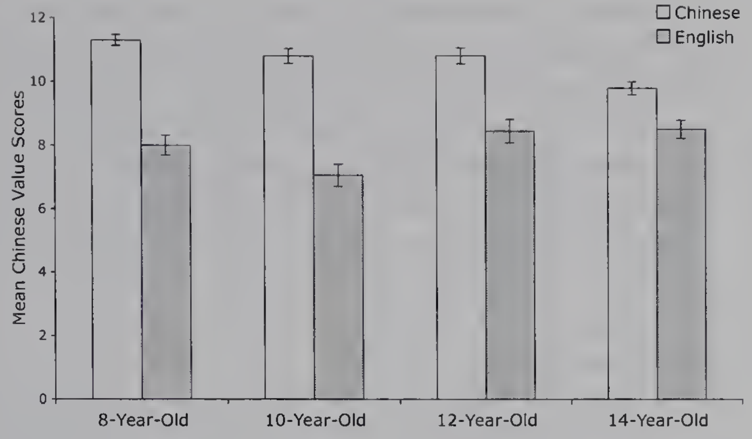


Figure 2. Mean Chinese value scores as a function of age and language of interview.

Endorsement of Cultural Values

Children on average scored 9.76 (*SD* = 2.47) for their endorsement of Chinese values. If responding in Chinese activates Chinese interdependent values and responding in English activates Western independent values, children in the Chinese interview condition should evidence higher agreement with Chinese values and children in the English interview condition should evidence higher agreement with Western values. Consistent with the hypothesis, children answering in Chinese ($M = 10.78$, $SD = 1.95$) reported higher agreement with Chinese interdependent values, and thus lower agreement with Western independent values, than did children answering in English ($M = 8.16$, $SD = 2.35$), $F(1, 109) = 43.29$, $p < .0001$, $\eta_p^2 = .28$. Figure 2 illustrates the mean Chinese value scores as a function of age and language of interview. In addition, girls ($M = 10.17$, $SD = 2.22$) reported greater endorsement of interdependent values than did boys ($M = 9.24$, $SD = 2.67$), $F(1, 109) = 7.81$, $p = .006$, $\eta_p^2 = .07$. There was no significant main effect or interaction related to age.

Self-Concept

Children on average provided 4.46 (*SD* = 8.67) self-descriptions. Inspection of the distribution showed that 3 children in the English interview condition were high-influence outliers, where the number of self-descriptions they provided (76, 34, and 32, respectively) were more than 3 *SD* above the group mean. When the 3 children were excluded from analysis, the average number of self-descriptions of the entire sample was 3.27 (*SD* = 3.57). A further ANOVA revealed that children in the English interview condition ($M = 4.60$, $SD = 3.77$) on average provided more self-descriptions than did children in the Chinese interview condition ($M = 2.42$, $SD = 3.20$), $F(1, 92) = 7.90$, $p = .006$, $\eta_p^2 = .08$. There were no significant effects pertaining to age or gender.

Next, the autonomous self score, that is, the ratio of personal-to-social self-descriptions ($M = 1.20$, $SD = 1.55$), was analyzed as a function of language, gender, and age. Because the ratio score controlled for variations in the total number of self-descriptions children provided, the 3 outlier children were included in the analysis. Consistent with the prediction, children interviewed in English ($M = 1.81$, $SD = 1.79$) exhibited a greater sense of autonomous self than did children interviewed in Chinese ($M = 0.78$, $SD = 1.20$), $F(1, 95) = 13.93$, $p = .0003$, $\eta_p^2 = .13$, regardless of gender or age. For example, a child interviewed in English responded that her story would include, "I like to play tennis . . . um, that I like to dance and um, I have a twin . . . that I like to do my homework and I like to swim . . . and I like, um, I have a favorite actor and singer." In comparison, a child interviewed in Chinese responded that in the story about her, she would "Write about my family; write about my mommy, daddy, and brother; my family lives in Jiulongtang." Figure 3 illustrates the mean autonomous self scores as a function of age and language of interview. No other main effect or interaction neared significance. Boys ($M = 1.37$, $SD = 1.68$) scored higher than girls ($M = 1.05$, $SD = 1.43$) on the autonomous self, although the difference was not significant.

Autobiographical Memory

The average length of children's memories was 37.62 (*SD* = 31.87). Analysis revealed a main effect of language, $F(1, 95) = 21.26$, $p < .0001$, $\eta_p^2 = .18$. As expected, children interviewed in English ($M = 51.32$, $SD = 27.12$) provided lengthier

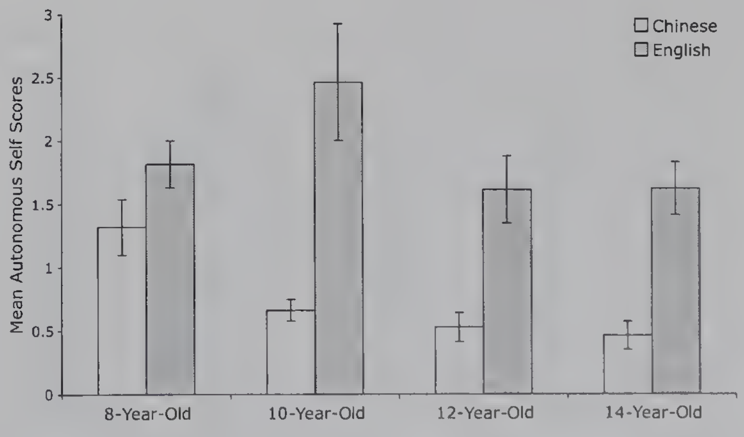


Figure 3. Mean autonomous self scores as a function of age and language of interview.

memories than did children interviewed in Chinese ($M = 28.28$, $SD = 31.67$). There were no significant effects pertaining to age or gender.

Analysis of the degree of self-focus in children's memory ($M = 1.10$, $SD = 0.65$), that is, the ratio of personal-to-social memory content descriptions, further revealed a main effect of language, $F(1, 95) = 5.25$, $p = .02$, $\eta_p^2 = .05$. Children interviewed in English ($M = 1.30$, $SD = 0.72$) provided more self-focused memories than did children interviewed in Chinese ($M = 0.97$, $SD = 0.58$), regardless of gender or age. Figure 4 illustrates the mean memory self-focus scores as a function of age and language of interview. There was no other significant main effect or interaction. Boys ($M = 1.12$, $SD = 0.70$) scored higher than girls ($M = 1.09$, $SD = 0.61$) on memory self-focus, although the difference was not significant.

The following two memory examples, both about a social conflict, further help illustrate the effects of language on memory retrieval. The first memory, provided by a child in the Chinese interview condition, focused on social interactions, group activities, and the roles of others, whereas the second memory, provided by a child in the English interview condition, highlighted the child's own roles, opinions, and perspectives.

> *Memory 1*: Once, my classmates and I were making a poster. One classmate wanted us to use his design. The rest of us didn't really like his idea. We didn't know what to do and argued with him. We all screamed, and the Head teacher had to come. In the end, even the Head teacher thought that he was wrong. So we started designing the poster all over again.

> *Memory 2*: Well, usually I argue with people because that's just sort of my nature. I like arguing. And recently I won this science competition thing at school and when I got the award I told one of my friends. Then he just didn't believe it so then I got a bit angry because he's supposed to believe it . . . well, basically, then I thought he was getting a bit selfish and stuff so I decided to just ignore him for a couple of days.

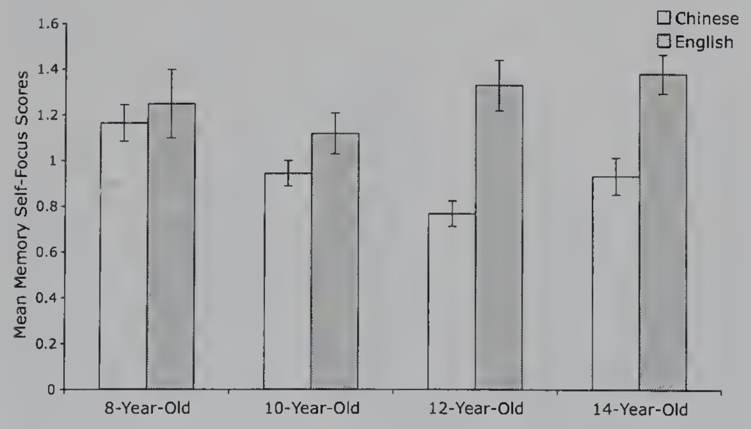


Figure 4. Mean memory self-focus scores as a function of age and language of interview.

Mediation Effects

To test the predicted mediation effects, two mediation analyses were conducted (Baron & Kenny, 1986; MacKinnon, Warsi, & Dwyer, 1995). Figure 5 illustrates the results from the analyses. The first analysis tested whether the effect of language on self-concept was mediated by the activation of cultural value systems (see Figure 5a). To establish the conditions for mediation, language (i.e., the predictor) was first regressed on Chinese value scores (i.e., the mediator) and autonomous self scores (i.e., the criterion), respectively (language was coded 1 for Chinese and 0 for English). The effects were significant in both cases, $t = 6.76$, $p < .0001$, and $t = -3.63$, $p = .0004$. A further regression confirmed that Chinese value scores had a significant effect on autonomous self scores, $t = -2.03$, $p = .04$, independent of language. Furthermore, when both language and Chinese value scores were simultaneously regressed on autonomous self scores, the independent contribution of language was substantially reduced, $t = -2.16$, $p = .03$,

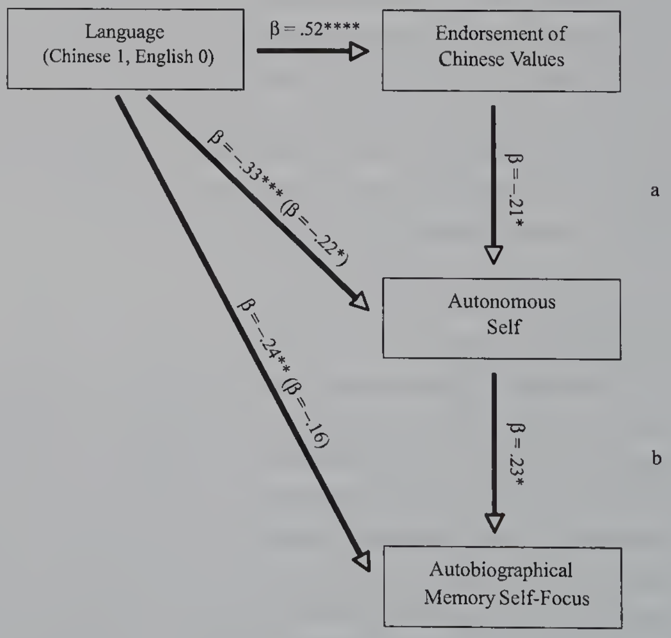


Figure 5. Mediation analyses.
Note. Language was coded 1 for Chinese and 0 for English. The coefficients in parentheses are from regressions when the mediator was included in the equation.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$. $^{****}p < .0001$.

whereby the standardized regression coefficient (β) of language decreased in both size and significance. The reduction was significant by the stringent Sobel (1982) test for mediation, $Z = -1.95$, $p = .05$. Thus, the activated cultural beliefs system is indeed a potent mediator for the effect of language on self-concept.

The next mediation model tested whether the working self-concept mediated the language effect on autobiographical memory (see Figure 5b). Language was a significant predictor of autonomous self scores, $t = -3.63$, $p = .0004$, and memory self-focus scores, $t = -2.62$, $p = .01$. Autonomous self scores were also related to memory self-focus scores, $t = 2.37$, $p = .02$, independent of language. Furthermore, the independent contribution of language to memory self-focus was reduced to a nonsignificant level by the effect of the mediator, the autonomous self, $t = -1.69$, $p = .09$. The reduction was significant by the Sobel (1982) test, $Z = -1.98$, $p = .05$. Thus, the activated autonomous self mediated the effect of language on memory self-focus.

Note that children were first tested for autobiographical recall, followed by the self-description task and then the cultural value questions. Were there any effects of the task sequence such that retrieving certain autobiographical information might lead children to adopt self-concepts consistent with such information, which in turn, might shape their endorsement of cultural values? To answer this question, we tested alternative models of the above analyses. We tested in the first model (Figure 5a) whether the autonomous self mediated the language effect on the endorsement of Chinese values instead. Then, we tested in the second model (Figure 5b) whether memory self-focus mediated the language effect on the autonomous self. In both alternative models, when language and the mediator were simultaneously regressed on the criterion, the effect of language was not significantly reduced compared with when language alone predicted the criterion, Sobel (1982) test $Z = 1.77$, *ns*, and $Z = -1.76$, *ns*, respectively. Thus, there were no mediations due to the task sequence. In addition, we tested a third model with language being the predictor, Chinese value scores being the mediator, and memory self-focus being the criterion. It was found that Chinese value scores did not relate to memory self-focus scores, $t = -0.94$, $p = .35$. The conditions for mediation were therefore not met (Baron & Kenny, 1986). Thus, the activation of culture belief systems did not have a direct effect on autobiographical memory, nor did it mediate the effect of language on memory.

Discussion

The present study is the first to test the effects of language on the cultural self among children. Two aspects of the self, namely, self-concept and autobiographical memory, were examined in bilingual children and adolescents from Hong Kong. Consistent with the predictions, across all age groups, children interviewed in English gave lengthier accounts about themselves and their personal experiences, and they focused more on unique personal attributes and qualities when describing themselves and talked more frequently about their own roles and perspectives in their autobiographical recall, compared with children interviewed in Chinese who provided less elaborate and more socially oriented self-descriptions and memory accounts. In addition, children responding in Chinese reported higher agreement with Chinese interdependent values and lower agreement with Western independent values than did those responding in English, regardless of age. Furthermore, the activation of cultural belief systems mediated the effect of language on self-concept, which, in turn, mediated the effect of language on autobiographical memory. Thus, linguistic usage (i.e., Chinese vs. English) triggered specific cultural belief systems (i.e., interdependence vs. independence), which further shaped the children's conceptions of self and autobiographical memory.

These findings suggest that bilingual children possess different modes of self associated with different languages. They support the theoretical notion that language as a representational medium of culture, represents, maintains, and further expresses the self through linguistic and conceptual configurations (e.g., syntax, grammar, vocabulary) and, perhaps more importantly, the associated (cultural) ways of thinking (Chiu et al., 2007; Holland & Quinn, 1987; Schrauf, 2000). Furthermore, as embedded practices, language learning and narrative interactions more generally take place in culturally organized communicative contexts that transmit to children not only linguistic and communicative skills but also cultural knowledge and competence (Fernald & Morikawa, 1993; Miller et al., 2007; Ochs, 1996). For instance, when sharing memories with their young children, European American mothers often focus on the child's roles and predilections in the past event and encourage the child to elaborate on his or her own perspectives. This child-centered, elaborative style facilitates children's detailed remembering of personal experiences that highlight their uniqueness and socialize children into an autonomous sense of self. In contrast, Chinese mothers often focus on shared activities, important others, and social norms and expect the child to answer pointed questions without elaboration. This group-oriented, pragmatic style helps children develop a sense of belonging and social compliance, and yet downplays the construction of elaborate personal stories and, further, a unique individual identity (Miller et al., 2007; Wang, Leichtman, & Davies, 2000). Through such joint activities, children gradually internalize culture-specific ways of defining themselves and remembering personal experiences (Wang, 2004, 2006a). Language is thus critical to the development and maintenance of a cultural self. For bilingual children, learning to speak two different languages equips them with multiple tools of constructing the self in relation to the sociocultural world. And developing different representations of the self in relation to different languages may further provide them with the flexibility to respond effectively to myriads of cultural and linguistic contexts that they transition in and out in everyday life (Ross et al., 2002; Schrauf, 2000).

The current findings are further provocative in terms of the specific mechanisms underlying the effect of language on the self. According to the cultural accommodation theory (Bond & Yang, 1982), the use of a language may create a ''response bias'' in the direction valued by the culture associated with that language, a process that may take place as either conscious self-presentation or unconscious automatic association. Although this theoretical view can account for the language effects on the children's agreement with Chinese versus Western independent values and their self-descriptions and memory recall, it does not predict the mediation effects. Alternatively, the constructivist view within a social-cognitive framework suggests that the use of a language may serve as a cue that makes salient and accessible the cultural belief system chronically associated with that language. The activated cultural belief system may further activate self-concepts congruent with the belief system. In turn, the currently operative working self-concepts may modulate retrieving of confirmatory autobiographical event information (Schrauf, 2000; Wang, 2008; Wang & Ross, 2005). The results from the mediation analyses lend support for this view, whereby responding in Chinese (vs. English) activated Chinese interdependent (vs. Western independent) values, which mediated, at least partially, the effect of language on the autonomous self. In turn, the autonomous self mediated the effect of language on the degree of self-focus in memory recall. These mediation effects were particularly notable given that the order of the tasks was arranged in a way that would work against the mediations. Indeed, the tests of alterative models showed that the mediations did not occur in opposite directions in spite of possible sequence effects among the tasks. It would be interesting to test in future studies if the expected mediation effects will become more potent when the task order is in favor of the mediations.

This ''chain of reactions'' has important implications for cross-cultural differences in self-concept and autobiographical memory (e.g., Cousins, 1989; Mullen, 1994; Rhee et al., 1995; Wang, 2001, 2004). In the daily life, language, no matter whether it is used privately in thinking or overtly in communication, may serve as a constant reminder of cultural knowledge and beliefs, which can further lead individuals to adopt self-concepts consistent with these knowledge structures. The culturally endorsed self-concepts, or cultural self-construals (Kağitçibaşi, 2007; Markus & Kitayama, 1991), may remain chronically active, salient, and accessible given the constant reinforcement of the linguistic-cultural context in which individuals reside. In turn, they may drive cognitive resources into the privileged processing and retrieving of autobiographical information that confirms the goals and motivations associated with the self-concepts (Conway & Pleydell-Pearce, 2000), thus giving rise to culture-specific ways of autobiographical remembering (Wang & Ross, 2007).

Notably, the activation of culture belief systems did not have a direct effect on autobiographical memory, nor did it mediate the effect of language on memory. This might seem counterintuitive given the cross-cultural findings that people in different cultures, upholding different cultural beliefs (e.g., independence vs. interdependence), tend to remember their autobiographical experiences in different ways (e.g., being self-focused vs. being socially oriented) consistent with their cultural beliefs (Han et al., 1998; Mullen, 1994; Wang, 2001, 2004; Wang & Leichtman, 2000). However, recent research examining specific mechanisms in individual social-cognitive processes that give rise to cultural differences suggests that culturally endorsed self-concepts function as a potent mediator or predictor in explaining cultural effects on autobiographical memory (Wang, 2006b, 2008; Wang & Ross, 2005). Similarly, the current study showed that the activation of culture belief systems indirectly influenced memory through the currently activated cultural self-concepts. These findings are consistent with general cognitive theories that emphasize the executive role of the conceptual self in modulating the constructive process of autobiographical remembering in line with the self's current goals and motivations (Conway, 2005; Conway & Pleydell-Pearce, 2000; Ross, 1989).

As expected, language effects were observed in children of all age groups. This, coupled with prior findings with adults (Kemmelmeier & Cheng, 2004; Marian & Kaushanskaya, 2004; Ross et al., 2002; Trafimow et al., 1997), suggests that the mechanisms underlying the language effects may operate similarly across age groups. On the other hand, unlike previous cross-cultural findings with preschool and early grade school children (Han et al., 1998; Wang, 2004), age effects and age-related interactions did not emerge in the current sample of children in middle childhood and early adolescence. This might be related to the nature of the tasks that had been previously used with younger children (Han et al., 1998; Wang, 2004, 2006b) and were therefore simple enough for older children and adolescents to give readily responses, thus eliminating age effects due to developmental differences in cognitive skills. Furthermore, the variables of interest might not differentiate among the children and adolescents of the limited age range of 8–14 years, who, at least in the current sample, showed similar levels of bilingual proficiency and might therefore have similar degrees of bicultural identification (Schrauf, 2000). Future research that includes children and adolescents of a wider age range is called for.

In line with the predictions, girls endorsed more strongly value systems emphasizing interdependence as opposed to independence than did boys, and their memories and self-descriptions appeared to be more socially oriented than those of boys, although the latter differences did not reach significance. These gender-related findings are consistent with the notion that gendered value orientations and self-knowledge emerge early though different family socialization practices and societal expectations (Cross & Madson, 1997; Fivush & Buckner, 2003). One important question to pursue further in this line of research is to examine differing developmental processes and consequences for boys and girls when gender norms and cultural norms coincide or clash.

The current findings have further implications for the development of the self in the immigration context. Many immigrant children and adolescents in the United States, for instance, are often exposed to contrasting cultural values and practices across various contexts of home, school, community, and the larger society (Mahalingam, 2006). Research has shown that developing a bicultural self, where immigrant individuals identify with both the host culture and their culture of origin, lends flexibility to cognition and behavior and is positively associated with psychological well-being (Kim & Omizo, 2006; Ryder, Alden, & Paulhus, 2000). In line with language socialization theories (Chiu et al., 2007; Miller et al., 2007; Ochs, 1996), findings from this study and previous studies with bilingual immigrant adults further suggest that bilingual proficiency may facilitate both the preservation of the ethnic culture and the adoption of the host culture, and may be a critical factor for the construction of a bicultural self among immigrants (Marian & Kaushanskaya, 2004; Ross et al., 2002; Schrauf, 2000). It should be noted, however, that different from the bilingual–bicultural children in Hong Kong, immigrant children may face unique obstacles as they strive to sustain their ethnic culture and language often without broader societal encouragement or recognition to do so, and this can make the development of a bicultural self an ever challenging task. Further longitudinal research is required to examine systematically how immigrant children come to develop a bicultural self in the process of acquiring the first and second languages and what important individual, familial, and societal factors are involved in this process.

George Miller once commented that "no general theory of psychology will be adequate if it does not take account of language" (Miller, 1990, p. 7). Applied to the self, it may follow that no theory of self-development will be adequate if it does not take account of language socialization. Our findings suggest that language is a powerful device for activating and reinforcing associated cultural constructs and plays a critical role in the construction, maintenance, and expression of a dynamic cultural self. In the process of becoming proficient speakers of their languages, children come to acquire cultural knowledge and culturally promoted self-views and are further socialized into competent members of their society.

References

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182.

Baumeister, R. F. (1998). The self. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 1, 4th ed., pp. 680–740). New York: McGraw-Hill.

Bond, M. H., & Yang, K. S. (1982). Ethnic affirmation versus cross-cultural accommodation: The variable impact of questionnaire language on Chinese bilinguals from Hong Kong. *Journal of Cross-Cultural Psychology, 13,* 169–185.

Bruner, J. (1990). *Acts of meaning*. Cambridge, MA: Harvard University Press.

Chiu, C.-Y., Leung, A. K.-Y., & Kwan, L. (2007). Language, cognition, and culture: Beyond the Whorfian hypothesis. In H. Kitayama & D. Cohen (Eds.), *Handbook of cultural psychology* (pp. 668–688). New York: Guilford.

Conway, M. A. (2005). Memory and the self. *Journal of Memory and Language, 53,* 594–628.

Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological Review, 107,* 261–288.

Cousins, S. D. (1989). Culture and self-perception in Japan and the United States. *Journal of Personality and Social Psychology, 56,* 124–131.

Cross, S. E., & Madson, L. (1997). Models of the self: Self-construals and gender. *Psychological Bulletin, 122,* 5–37.

Deeds, O., Stewart, S. M., Bond, M. H., & Westrick, J. (1998). Adolescents between cultures: Values and autonomy expectations in an international school setting. *School Psychology International, 19,* 61–78.

Feldman, S. S., & Rosenthal, D. A. (1991). Age expectations of behavioural autonomy in Hong Kong, Australian and American youth: The influence of family variables and adolescents' values. *International Journal of Psychology, 26,* 1–23.

Fernald, A., & Morikawa, H. (1993). Common themes and cultural variations in Japanese and American mothers' speech to infants. *Child Development, 64,* 637–656.

Fivush, R., & Buckner, J. P. (2003). Creating gender and identity through autobiographical narratives. In R. Fivush & C. Haden (Eds.), *Autobiographical memory and the construction of a narrative self: Developmental and cultural perspectives* (pp. 149–167). Hillsdale, NJ: Erlbaum.

Fivush, R., Haden, C., & Adam, S. (1995). Structure and coherence of preschoolers' personal narratives over time: Implications for childhood amnesia. *Journal of Experimental Child Psychology, 60,* 32–56.

Fivush, R., & Nelson, K. (2004). Culture and language in the emergence of autobiographical memory. *Psychological Science, 15,* 573–577.

Han, J. J., Leichtman, M. D., & Wang, Q. (1998). Autobiographical memory in Korean, Chinese, and American children. *Developmental Psychology, 34,* 701–713.

Hart, D., Lucca-Irizarry, N., & Damon, W. (1986). The development of self-understanding in Puerto Rico and the United States. *Journal of Early Adolescence, 6,* 293–304.

Harter, S. (1999). *The construction of the self: A developmental perspective*. New York: Guilford.

Holland, D., & Quinn, N. (1987). *Cultural models in language and thought*. New York: Cambridge University Press.

Hong, Y.-y., Morris, M. W., Chiu, C.-y., & Benet-Martinez, V. (2000). Multicultural minds: A dynamic constructivist approach to culture and cognition. *American Psychologist, 55,* 709–720.

Kağitçibaşi, C. (2007). *Family, self, and human development across cultures* (2nd ed.). Mahwah, NJ: Erlbaum.

Keller, A., Ford, L. H., & Meacham, J. A. (1978). Dimensions of self-concept in preschool children. *Developmental Psychology, 14,* 483–489.

Kemmelmeier, M., & Cheng, B.-Y. (2004). Language and self construal priming; A replication and extension in a Hong Kong sample. *Journal of Cross-Cultural Psychology, 35,* 705–712.

Kim, B. S. K., & Omizo, M. M. (2006). Behavioral acculturation and enculturation and psychological functioning among Asian American college students. *Cultural Diversity and Ethnic Minority Psychology, 12,* 245–258.

MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research, 30,* 41–62.

Mahalingam, R. (Ed.). (2006). *Cultural psychology of immigrants*. Mahwah, NJ: Erlbaum.

Marian, V., & Kaushanskaya, M. (2004). Self-construal and emotion in bicultural bilinguas. *Journal of Memory and Language, 51,* 190–201.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98,* 224–253.

Miller, G. A. (1990). The place of language in a science psychology. *Psychological Science, 1,* 7–14.

Miller, P. J., Fung, H., & Koven, M. (2007). Narrative reverberations: How participation in narrative practices co-creates persons and cultures. In H. Kitayama & D. Cohen (Eds.), *Handbook of cultural psychology* (pp. 595–614). New York: Guilford.

Mullen, M. K. (1994). Earliest recollections of childhood: A demographic analysis. *Cognition, 52,* 55–79.

Neisser, U. (1988). Five kinds of self-knowledge. *Philosophical Psychology, 1,* 35–59.

Nelson, K. (2001). Language and the self: From the "Experiencing I" to the "Continuing Me." In C. Moore & K. Lemmon (Eds.), *The self in time: Developmental perspectives* (pp. 15–34). Mahwah, NJ: Erlbaum.

Ochs, E. (1996). Linguistic resources for socialising humanity. In J. Gumperz & S. Levinson (Eds.), *Rethinking linguistic reality* (pp. 407–437). Cambridge, UK: Cambridge University Press.

Rhee, E., Uleman, J. S., Lee, H. K., & Roman, R. J. (1995). Spontaneous self-descriptions and ethnic identities in individualistic and collectivistic cultures. *Journal of Personality and Social Psychology, 69,* 142–152.

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis*. New York: McGraw-Hill.

Ross, M. (1989). Relation of implicit theories to the construction of personal histories. *Psychological Review, 96*, 341–357.

Ross, M., Xun, W. Q. E., & Wilson, A. E. (2002). Language and the bicultural self. *Personality and Social Psychology Bulletin, 28*, 1040–1050.

Ryder, A. G., Alden, L. E., & Paulhus, D. L. (2000). Is acculturation unidimensional or bidimensional? A head-to-head comparison in the prediction of personality, self-identity, and adjustment. *Journal of Personality and Social Psychology, 79*, 49–65.

Schrauf, R. W. (2000). Bilingual autobiographical memory: Experimental studies and clinical cases. *Culture & Psychology, 6*, 387–417.

Singelis, T. M. (1994). The measurement of independent and interdependent self-construals. *Personality and Social Psychology Bulletin, 20*, 580–591.

Sobel, M. E. (1982). Asymptotic intervals for indirect effects in structural equations models. In S. Leinhart (Ed.), *Sociological methodology 1982* (pp. 290–312). San Francisco: Jossey-Bass.

Stigler, J., Smith, S., & Mao, L. (1985). The self-perception of competence by Chinese children. *Child Development, 56*, 1259–1270.

Trafimow, D., Silverman, E. S., Fan, R. M., & Law, J. S. F. (1997). The effects of language and priming on the relative accessibility of the private self and the collective self. *Journal of Cross-Cultural Psychology, 28*, 107–123.

Trafimow, D., Triandis, H. C., & Goto, S. G. (1991). Some tests of the distinction between the private self and the collective self. *Journal of Personality and Social Psychology, 60*, 649–655.

Tung, P., Lam, R., & Tsang, W. K. (1997). English as a medium of instruction in post-1997 Hong Kong: What students, teachers, and parents think. *Journal of Pragmatics, 28*, 441–459.

Umaña-Taylor, A. J., Yazedjian, A., & Bámaca-Gómez, M. Y. (2004). Developing the Ethnic Identity Scale using Eriksonian and social identity perspectives. *Identity: An International Journal of Theory and Research, 4*, 9–38.

Valsiner, J. (2001). Process structure of semiotic mediation in human development. *Human Development, 44*, 84–97.

Wang, Q. (2001). Cultural effects on adults' earliest childhood recollection and self-description: Implications for the relation between memory and the self. *Journal of Personality and Social Psychology, 81*, 220–233.

Wang, Q. (2004). The emergence of cultural self-constructs: Autobiographical memory and self-description in European American and Chinese children. *Developmental Psychology, 40*, 3–15.

Wang, Q. (2006a). Culture and the development of self-knowledge. *Current Directions in Psychological Science, 15*, 182–187.

Wang, Q. (2006b). Relations of maternal style and child self-concept to autobiographical memories in Chinese, Chinese immigrant, and European American 3-year-olds. *Child Development, 77*, 1794–1809.

Wang, Q. (2008). Being American, being Asian: The bicultural self and autobiographical memory in Asian Americans. *Cognition, 107*, 743–751.

Wang, Q., & Leichtman, M. D. (2000). Same beginnings, different stories: A comparison of American and Chinese children's narratives. *Child Development, 71*, 1329–1346.

Wang, Q., Leichtman, M. D., & Davies, K. (2000). Sharing memories and telling stories: American and Chinese mothers and their 3-year-olds. *Memory, 8*, 159–177.

Wang, Q., & Ross, M. (2005). What we remember and what we tell: The effects of culture and self-priming on memory representations and narratives. *Memory, 13*, 594–606.

Wang, Q., & Ross, M. (2007). Culture and memory. In H. Kitayama & D. Cohen (Eds.), *Handbook of cultural psychology* (pp. 645–667). New York: Guilford.

Child Development, March/April 2010, Volume 81, Number 2, Pages 568–580

When Do Children Exhibit a "Yes" Bias?

Mako Okanda and Shoji Itakura
Kyoto University

This study investigated whether one hundred and thirty-five 3- to 6-year-old children exhibit a yes bias to various yes–no questions and whether their knowledge status affects the production of a yes bias. Three-year-olds exhibited a yes bias to all yes–no questions such as *preference-object* and *knowledge-object* questions pertaining to objects, and *knowledge-face* questions pertaining to facial expressions. Four-year-olds tended to say "yes" only to knowledge-object questions. Five-year-olds did not show any strong response tendency. Six-year-olds exhibited a *nay-saying* bias to knowledge-face questions. Also, 3-year-olds could indicate the correct option when asked questions with 2 response options. It suggested that 3-year-olds tended to inappropriately say "yes" to yes–no questions, although they knew the answers to the questions. The mechanism of a yes bias was discussed.

A large number of studies in developmental psychology often use questioning of children in their research methodology. Researchers question children in experimental tasks in order to obtain conclusions about the children's social and cognitive development (e.g., Fritzley & Lee, 2003; Waterman, Blades, & Spencer, 2004); however, children sometimes exhibit a response bias to questions. A response bias is seen when children tend to give inappropriate responses by providing the same answer such as saying "yes" to yes–no questions (i.e., a yes bias).

A response bias is most often observed when children are asked closed questions like forced-choice and yes–no questions (e.g., Flavell, Flavell, & Green, 1983; Flavell, Green, & Flavell, 1986; Flavell, Zhang, Zoui, Dong, & Qi, 1983; Gopnik, Sobel, Schulz, & Glymour, 2001; Siegal & Surian, 2007; Steffensen, 1978), with yes–no questions being the most problematic type for preschoolers (see also Fritzley & Lee, 2003). In fact, a yes bias was stronger when 3- to 5-year-olds were asked yes–no questions as compared to open-ended questions (i.e., wh questions; Peterson, Dowden, & Tobin, 1999) or forced-choice questions (Peterson & Grant, 2001). Five- to 9-year-old English-speaking children could say "I don't know" to unanswerable wh questions, whereas they were reluctant to say it and attempted to provide "yes" or "no" responses to unanswerable yes–no questions (Waterman, Blades, & Spencer, 2001; Waterman et al., 2004). Similarly, 5-year-olds tended to say "yes" to bizarre yes–no questions such as "Is milk bigger than water?" (Hughes & Grieve, 1980).

Given the possibility that children exhibit a yes bias to yes–no questions, several studies investigated children's yes bias in simulated interviews about staged or experienced events (Peterson & Grant, 2001; Peterson et al., 1999), in simple experimental situations (Fritzley & Lee, 2003; Okanda & Itakura, 2008), and in daily conversations (Okanda & Itakura, 2007; see also Steffensen, 1978). Fritzley and Lee (2003) found that 2- and 3-year-olds exhibited a yes bias to simple yes–no questions pertaining to the properties and functions of objects. Okanda and Itakura (2008) replicated the study and found that 2- to 4 year-old Japanese and Vietnamese children exhibited a yes bias to the same type of yes–no questions.

This research was supported by the research fellowship of the Japan Society for the Promotion of Science (JSPS) for young scientists to Mako Okanda, by the grants from the JSPS (20330150) and the Nissan Science Foundation to Shoji Itakura, and by the Global COE Program "Revitalizing Education for Dynamic Hearts and Minds," MEXT, Japan. We are grateful to the teachers in the kindergartens and the nursery schools, the parents of the children, and especially to the children who participated in this study. We are also grateful to the two anonymous reviewers and to Kazuo Fujita, Makoto Takahashi, Yusuke Moriguchi, Tara Callaghan, and Kang Lee for their constructive and valuable comments on experiments and/or drafts.

Correspondence concerning this article should be addressed to Mako Okanda, Department of Psychology, Graduate School of Letters, Kyoto University, Yoshida-honmachi, Sakyo, Kyoto 606 8501, Japan. Electronic mail may be sent to mako-okanda@l03.mbox.media.kyoto-u.ac.jp.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8102-0013

A yes bias appears to be a common phenomenon among young preschoolers who are living in different countries and speak different languages. Three main reasons have been advanced to account for a yes bias. The first relates more to perceived social factors, and the second and third to cognitive development.

First, children might feel social pressures that they have to provide an answer when questioned by adults (Cassel, Roebers, & Bjorklund, 1996; Moston, 1990) because they might believe that (a) adults prefer a "yes" response (Brady, Poole, Warren, & Jones, 1999), (b) it would be in agreement with adults (Peterson & Biggs, 1997), and (c) it would display compliance (Peterson & Biggs, 1997) with the interviewer's intentions, especially when he or she is not a family member (Krähenbühl & Blades, 2006; Zajac, Gross, & Hayne, 2003; Zajac & Hayne, 2003). In addition to these social pressures, children might be likely to affirm interviewers' intentions because yes–no questions are highly suggestive (Ceci & Bruck, 1993; Peterson et al., 1999). The social pressure research typically examines children's answers to experimenters, and thus, the question arises as to whether children provide inappropriate answers to yes–no questions because the interviewers are unfamiliar adults. However, Okanda and Itakura (2007) found that 2- and 3-year-olds exhibited a yes bias to their own mothers. The issue of the social pressure of compliance still remains in this case, but it is clear that a stranger effect does not explain fully why children show a yes bias.

The second and third possibilities are that young preschoolers may be unable to appropriately answer questions due to cognitive factors that impact their ability to understand conversations or pragmatics with the experimenter, or their inhibitory control. Related to the second possibility, Lee and Eskritt (1999) suggested that understanding the contexts of questions, such as what an interviewer would like to know, is considerably important in order to appropriately answer questions. Children may have learned that saying "yes" is better than saying "no" in day-to-day questions (e.g., "Could you close the door?"), and thus, they tend to say "yes" to yes–no questions without fully understanding the contexts of the questions. Similarly, Fritzley and Lee (2003) suggested that children's yes bias and conversational or pragmatic understanding might be related. The third possibility suggests that a "yes" response could be a dominant response that is difficult to inhibit for young preschoolers (Okanda & Itakura, 2007). Moriguchi, Okanda, and Itakura (2008) indicated that the dominance of "yes" responses may be achieved simply because adults rarely ask young children questions for which the answers are "no" and so children may have more experience in saying "yes" than in saying "no." Evidence that supports the idea that these are mitigating cognitive factors suggests that children who have lesser verbal ability and inhibitory control displayed a stronger yes bias (Moriguchi et al., 2008).

In this study, we hypothesized that cognitive factors play a role in a yes bias and examined the following three issues that have not been investigated thus far in previous studies: (a) Do children exhibit a yes bias when asked yes–no questions about their preferences? (b) Do children display a yes bias to yes–no questions pertaining to knowledge of objects as well as to different domains, such as questions about knowledge of facial expressions? (c) Do children exhibit a yes bias even when they actually know the answer?

The first research question investigates whether children exhibit a yes bias to yes–no questions pertaining to their preference. Previous research only considered questions that pertain to what the children knew, did, or witnessed; this is probably because most of the previous studies were concerned with the accuracy of children's responses to questions in forensic interviews. For example, Peterson et al. (1999) questioned children after they were shown staged events. Waterman et al. (2001) narrated short stories and questioned children about these stories. Although Fritzley and Lee (2003) did not employ a simulated forensic interview style, they asked questions only pertaining to the properties and functions of objects. Thus, the questions used in the previous studies were all knowledge questions that demanded objective judgments and that evaluated children's knowledge.

In Study 1, we asked children "preference" yes–no questions pertaining to objects (e.g., "Do you like this?" or "Do you dislike this?"; i.e., *preference-object yes–no questions*) that demanded binary but subjective decision making. Such questions are clearly different from the questions asked in the previous studies pertaining to the properties and functions of objects (i.e., *knowledge-object yes–no questions*) because here children were allowed to answer "yes" or "no" depending on their preferences, and there were no particular correct answers. More importantly, children should have more experiences with preference than knowledge questions in daily conversation, and thus, the former may be easier for children to answer appropriately than the latter. It is worth comparing these two types of questions. In addition, although the word *dislike* is an unfamiliar word in English, the word meaning *dislike* in Japanese, *kirai*, is a very familiar word for Japanese preschoolers. We chose *dislike* rather than *hate* which may be familiar to English-speaking preschoolers, because the English *hate* is too extreme to use for the Japanese *kirai*.

The second research question investigates whether children exhibit a yes bias to yes–no questions pertaining to social entities. In Study 2, we asked children simple yes–no questions pertaining to facial expressions (i.e., *knowledge-face yes–no questions*) in order to compare their response tendencies to knowledge-based questions. Fritzley and Lee (2003) asked children simple questions about objects, but they did not ask questions about social entities. Although other studies have questioned children about events, environments, and people, these studies were simulated forensic interviews with a delay between the events and interviews that might introduce memory issues (e.g., Peterson & Biggs, 1997; Peterson & Grant, 2001; Peterson et al., 1999). Questions about facial expressions are a special type of knowledge question, one that taps children's social emotional knowledge.

The last issue we considered was to examine whether children's knowledge status influences the production of a yes bias. In Study 3, we confirmed what children's knowledge status was and then asked children choice questions that had two options (i.e., ''Which is X?'' and ''Which is for X?''). We presented children with two objects or cards that had been used in the first two studies in this study, as well as items from the study conducted by Okanda and Itakura (2007). We investigated the possibility that children might say ''yes'' when asked yes–no questions, despite knowing the answer.

In addition, we examined whether Japanese children give ''I don't know'' and ''no answer'' responses in Study 1 and Study 2 because Okanda and Itakura (2008) reported that Asian children, in contrast to North American children, do show these responses to yes–no questions.

Study 1

In this study, we investigated whether children show a yes bias to preference-object questions that have no correct answer, framed in a social context to examine children's attitudes. Children's responses to the preference-object questions were compared with those to knowledge-object questions that have correct answers.

Method

Participants

Participants of this study were 61 Japanese children: twenty 3-year-olds (M = 43.3 months, SD = 2.70, range = 37–47 months; 8 boys and 12 girls), twenty 4-year-olds (M = 54.1 months, SD = 3.69, range = 48–59 months; 9 boys and 11 girls), and twenty-one 5- to 6-year-olds (M = 72.4 months, SD = 3.76, range = 64–79 months; 11 boys and 10 girls). An additional 3-year-old child was investigated but excluded from further analysis because she refused to participate in the experiment. The children were selected from kindergartens and nursery schools in Kyoto and Kobe, Japan.

Materials and Procedures

We followed the procedure and materials used by Fritzley and Lee (2003) and Okanda and Itakura (2008). Objects used in the present study were identical to those used by Okanda and Itakura (2008). The familiar objects included a blue cup, a red apple, and a picture book, and the unfamiliar objects included a plastic coffee filter, a shoehorn, and a CPU. Children's object familiarity to these items was confirmed in Okanda and Itakura (2008). The familiar objects were known to be familiar, and the unfamiliar objects were known to be unfamiliar for Japanese 2- to 5-year-old children.

In the experiment, each child was interviewed individually in a classroom at the nursery school in which he or she was studying. Each child was asked 36 questions, with six questions pertaining to each object (the list of the questions asked is provided in Appendix A). Four of the six questions were knowledge-object yes–no questions that inquired about the properties and functions of objects, and the other two were preference-object yes–no questions that asked whether the child liked or disliked the object. Two of the four knowledge questions were yes questions (i.e., the correct answer was ''yes''), and the other two questions were no questions (i.e., the correct answer was ''no''). The two preference questions had no correct answer, and each child could provide any answer that he or she felt to be appropriate. For every object, each child was asked both like and dislike questions (i.e., ''Do you like this? ''Do you dislike this?") to confirm whether children appropriately and consistently showed their preferences. That is, if the children liked the object, they were required to answer "yes" to the like question and "no" to the dislike question.

Half of the children were first asked questions about a familiar object. An experimenter randomly chose one of three familiar objects as the first target object and subsequently asked six questions. The six questions, both preference and knowledge questions with regard to each object, were presented to each child in a random manner. The experimenter then randomly chose one of three unfamiliar objects as the second target object and asked six questions. The following target objects alternated between the remaining familiar and unfamiliar objects. For the other half of the children, the first target object was an unfamiliar object, and the subsequent target objects were alternately chosen from familiar and unfamiliar objects. Children did not always need to know the exact names and functions of the objects to answer the yes–no questions. For example, some questions asked about color or shape of the objects. The experimenters explained the purpose and method of the study to the directors of the nursery schools prior to the experiments.

Scoring

Knowledge response bias scores. We modified Fritzley and Lee's (2003) scoring method to examine the response bias score for the knowledge yes–no questions for each child (a knowledge response bias score). First, for all yes questions the children scored 1 if their response was "yes" (a correct answer) and −1 if their response was "no" (an incorrect answer). The scores were added to obtain a yes score. Similarly, for all no questions, the children scored 1 for all "no" responses (a correct answer) and −1 for all "yes" responses (an incorrect answer). Thereafter, the scores were added to obtain a no score. Both "no answer" (when they were posed and said nothing) and "I don't know" responses were scored as zero. To derive a proportional yes score, the yes score was divided by the total number of yes questions. A proportional no score was derived in a similar manner. Subsequently, the proportional no score was subtracted from the proportional yes score and a knowledge response bias score was obtained. The score range was 1 to −1, with a positive response bias score indicating a yes bias and a negative response bias score indicating a nay-saying bias (a response tendency to say "no"). The knowledge response bias score for a child who did not have a response bias would be zero.

Preference response bias scores. The calculation of a response bias score for the preference yes–no questions (a preference response bias score) was modified based on the calculation of the knowledge response bias scores. In order to examine whether each child's responses corresponded with each other in both the like and dislike questions for each object, the children were given 1 for every "yes" response and −1 for every "no" response. Thereafter, the mean scores were considered as the preference response bias score. The preference response bias score for a child who did not have a response bias was zero (i.e., a child who was able to answer "yes" to like questions and "no" to dislike questions, or vice versa). A positive preference response bias score implied that the child tended to respond with a "yes" to both the like and dislike questions, and a negative preference response bias score suggested that a child tended to respond with a "no" to both questions. Both the knowledge and preference response bias scores were separately calculated for familiar and unfamiliar objects.

Results

Figure 1 shows the children's response bias scores in each age group. We conducted a preliminary analysis on the 5- and 6-year-olds because the age range of the 5- to 6-year-olds was 64–79 months and the 6-year-olds were actually younger

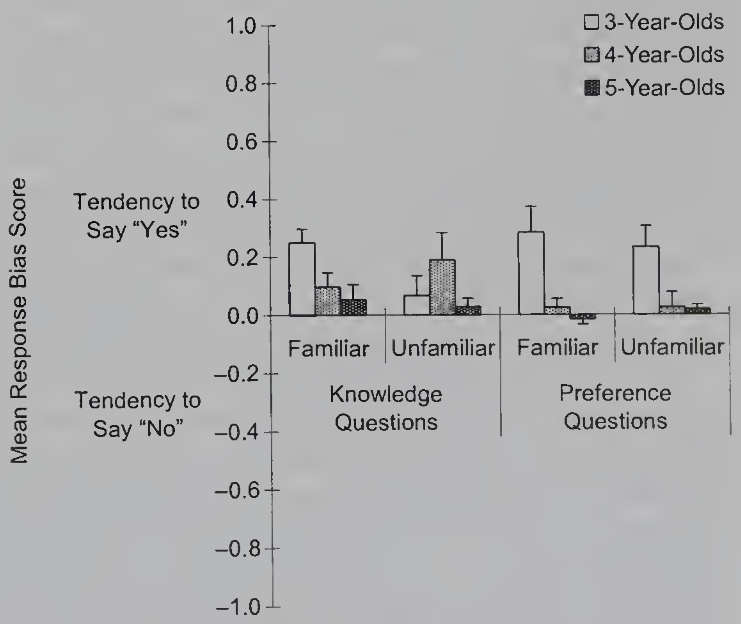


Figure 1. Children's mean knowledge response bias scores and mean preference response bias scores on knowledge and preference questions.
Note. T-bars represent standard errors.

6-year-olds. There was no main effect of age between 5- and 6-year-old children; hence, we combined them into one group. Subsequently, a 3 × 2 × 2 mixed-design analysis of variance (ANOVA; age: 3 years, 4 years, 5–6 years; familiarity: familiar, unfamiliar; question type: knowledge, preference) was conducted. The main effect of age was significant, $F(2, 58) = 7.94$, $p < .01$, partial $\eta^2 = .22$. Post hoc Tukey honestly significant difference (HSD) tests revealed that 3-year-olds showed a significantly stronger yes bias than did the 4- ($p < .05$) and 5-year-olds ($p < .01$). Furthermore, the Age × Familiarity interaction was significant, $F(2, 58) = 3.22$, $p < .05$, partial $\eta^2 = .10$, and the Age × Question Type interaction was marginally significant, $F(2, 58) = 2.93$, $p = .061$, partial $\eta^2 = .09$.

One-sample t tests were conducted to confirm these interactions and to examine whether the response bias scores were significantly above or below zero. The mean knowledge and preference response bias scores of each age group within each condition was compared to a score of zero (a zero implied that there was no response bias). With regard to the knowledge questions, 3-year-olds exhibited a strong yes bias with respect to familiar condition, $t(19) = 5.31$, $p < .01$, and 4-year-olds tended to display a yes bias for both familiar and unfamiliar items, $t(19) = 1.91$, $p < .10$, and $t(19) = 2.01$, $p < .10$, respectively. In addition, the one-sample t tests indicated that the mean knowledge response bias score of 5- to 6-year-olds was significantly higher than zero with respect to familiar condition, $t(20) = 2.80$, $p < .05$, although the raw knowledge response bias score was almost zero ($M = .053$). The knowledge response bias scores of 5- to 6-year-olds were significantly lower than those of 3-year-olds. Japanese 5- to 6-year-olds could have a yes bias, but the tendency was very weak. With regard to the preference questions, only 3-year-olds exhibited a strong yes bias with respect to both familiar and unfamiliar conditions, $t(19) = 3.21$, $p < .01$, and $t(19) = 3.19$, $p < .01$, respectively.

In addition, the frequency of "I don't know" and "no answer" response was also calculated. For the knowledge yes–no questions, the children were more likely to provide other responses ("I don't know") or not respond when they were asked unfamiliar questions (12 times) than when they were asked familiar questions (6 times). For the preference yes–no questions, all the children responded to the questions. There was no difference in the "I don't know" responses for familiar and unfamiliar objects; the frequency of the "I don't know" responses was 9 versus 11, respectively.

Discussion

In this study, we investigated whether 3- to 6-year-old Japanese preschoolers exhibit a yes bias to preference-object and knowledge-object yes–no questions. The results revealed that 3-year-olds had strong tendency to exhibit a yes bias to both preference and knowledge questions. Four-year-olds were able to appropriately answer preference questions but showed a yes bias to knowledge questions. Five- to 6-year-olds did not show a response bias to preference questions but showed very weak tendency to say "yes" to knowledge questions regarding familiar objects.

First, the results with regard to the knowledge questions in this study were mostly consistent with those obtained in a similar study by Okanda and Itakura (2008). Second, the most important finding in this study was that 3-year-olds exhibited a yes bias to both knowledge and preference questions, but 4- to 6-year-olds displayed a yes bias only to knowledge yes–no questions. Knowledge and preference questions might be equally difficult for 3-year-olds; however, knowledge questions might be more difficult than preference questions for 4- to 6-year-olds. Hence, the older children were likely to display a yes bias to difficult yes–no questions. This result may indicate that 3-year-olds show a yes bias automatically regardless of the question type, but older children may have different reasons for responding with a "yes" in certain questions. It is possible that some social factors may affect older children's yes bias; however, a weak yes bias in 5- to 6-year-olds in this study must be treated with caution because the previous study did not report such a tendency (Okanda & Itakura, 2008).

Third, object familiarity did not influence the children's yes bias with respect to preference questions. In contrast, it might be important with regard to the knowledge question because the children's yes bias differed between questions regarding familiar and unfamiliar objects. One possible reason why object familiarity influenced a yes bias only for knowledge questions is that the children were in a test situation where their knowledge was being examined. In this kind of a situation, the children were sensitive to whether they were knowledgeable or ignorant about the questions and thus it affected their answers. Our result indicates that knowledge and preference questions differ in terms of their inherent characteristics.

Finally, Japanese children in this study were likely to show a stronger yes bias to familiar objects than to unfamiliar objects, as Okanda and Itakura

(2008) reported. This tendency was exactly opposite to that obtained by Fritzley and Lee (2003) when they examined North American children. However, this tendency was not robustly observed in this study, and we need to further examine the object familiarity effect and cross-cultural differences in future research.

In this study, we found that 3-year-olds display a yes bias to both knowledge and preference questions, but older children exhibited a weak yes bias only to knowledge questions. In Study 2, we examined whether children exhibit a yes bias to knowledge questions from the social domain.

Study 2

In this study, we investigated whether children exhibit a yes bias to yes–no questions pertaining to the facial expressions of schematic faces. It is important to note that while a question about facial expressions is a type of knowledge question, it pertains to social entities. Children's responses to knowledge-face questions were compared with their responses to knowledge-object questions.

Method

Participants

Participants of this study included 74 Japanese children, none of whom participated in Study 1: twenty-seven 3-year-old children (M = 41.1 months, SD = 3.57, range = 36–47 months; 14 boys and 13 girls), twenty-two 4-year-old children (M = 54.1 months, SD = 3.24, range = 48–59 months; 15 boys and 7 girls), twelve 5-year-old children (M = 67.1 months, SD = 3.53, range = 60–71 months; 9 boys and 3 girls), and thirteen 6-year-old children (M = 76.5 months, SD = 3.73, range = 72–83 months, 9 boys and 4 girls). The children were selected from kindergartens and nursery schools in Kyoto and Ashiya, Japan.

Materials and Procedures

For the object condition, we used four familiar objects. Three objects were exactly the same as those used in Study 1 (i.e., a blue cup, a red apple, and a picture book), and the fourth object was a spoon. For the face condition, four schematic faces were used. Two cards were used for each face: One showed the face smiling and the other showed it crying; a total of eight cards were used.

In the experiment, each child was tested individually. A child was asked 32 questions: 16 object questions and 16 face questions. For the object condition, two of four questions were yes questions (i.e., questions whose correct answers were "yes") and the other two questions were no questions (i.e., questions whose correct answers were "no"). For the face condition, each card had two questions: a yes question and a no question (see Appendix B).

Half of the children were first asked object questions, and the remaining half of the children were first asked face questions. For the former group, an experimenter randomly chose one of the four objects and asked the children four questions. The order of four questions was randomized. Thereafter, the experimenter randomly chose objects sequentially from the remaining objects and asked four questions per object accordingly. For the latter group, the experimenter randomly chose one of the eight facial expression cards, showed it to the child, and asked two questions in random order. The experimenters explained the purpose and method of the study to the directors of the nursery schools prior to the experiments.

Results

The response bias scores of the object and face conditions were calculated separately. The scoring method used in this study was exactly the same as that used in Study 1. Age range of the 5- and 6-year-olds in this study was 60–83 months. A preliminary analysis revealed that the 5- and 6-year-olds exhibited extremely different response bias patterns; hence, we analyzed them under different age groups. We conducted a 4 × 2 × 2 mixed-design ANOVA (age: 3 years, 4 years, 5 years, 6 years; entity: object, face; order: object questions first, face questions first) to examine whether there was an order effect between the groups who received the object versus face condition first. There was no order effect, $F(1, 66) = .06$, $p = .81$, partial $\eta^2 = .01$, and no interaction between order and entity was revealed, $F(3, 66) = 1.05$, $p = .36$, partial $\eta^2 = .05$.

A 4 × 2 mixed-design ANOVA (age: 3 years, 4 years, 5 years, 6 years; entity: object, face) was then conducted on the children's response bias scores (see Figure 2). The main effect of age was significant, $F(3, 70) = 10.43$, $p < .01$, partial $\eta^2 = .31$. Post hoc Tukey HSD tests revealed that the 3-year-olds' mean response bias scores were significantly higher than that of 4- ($p < .01$), 5- ($p < .01$), and 6-year-olds ($p < .01$). Furthermore, the main effect of entity was also significant, $F(1, 70) = 22.77$, $p < .01$,

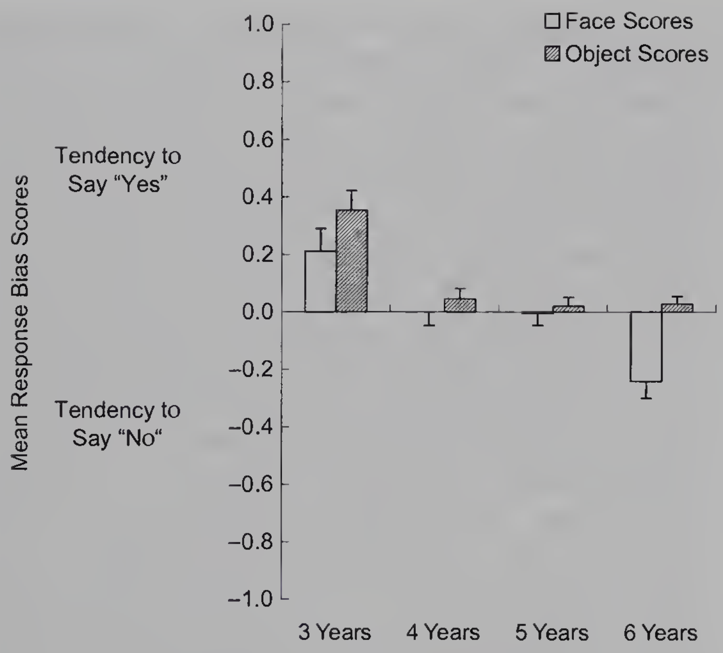


Figure 2. Children's mean knowledge response bias scores on object and face questions.
Note. T-bars represent standard errors.

partial $\eta^2 = .25$. The children exhibited a stronger yes bias to object questions than they did to face questions. No interaction between age and entity was revealed, $F(3, 70) = 2.10$, $p = .11$, partial $\eta^2 = .08$.

One-sample t tests were conducted to confirm whether response bias scores were significantly above or below zero. The mean response bias score of each age group and each condition was compared to a score of zero, which indicated no response bias. Three-year-olds exhibited a strong yes bias in the object condition, $t(26) = 5.25$, $p < .01$, and the face condition, $t(26) = 2.63$, $p < .05$, whereas 6-year-olds displayed a nay-saying bias in the face condition, $t(10) = -4.06$, $p < .01$.

Neither face questions nor object questions elicited a large number of "I don't know" and "no answer" responses. Only one 4-year-old child displayed a high frequency of not responding to the questions.

Discussion

Study 2 investigated whether the type of question (object vs. face) influenced children's yes bias. As in Study 1, the results confirmed that 3-year-olds exhibited a strong yes bias to various yes–no questions and showed further that the tendency to say "yes" was stronger to object questions than to face questions. In addition, 6-year-olds displayed a nay-saying bias to face questions.

First, 3-year-olds exhibited a strong yes bias to both knowledge-objects and knowledge-face questions, and their yes bias was the strongest among all age groups. This suggests that 3-year-olds display a yes bias to yes–no questions regardless of the entities to which the questions pertain. It is possible that 3-year-olds automatically exhibit a yes bias when asked yes–no questions.

Second, entities to which the questions pertained possibly influenced the strength of children's response bias. Overall, children in Study 2 said "yes" more frequently to object questions than they did to face questions. This result is partly consistent with the result of Study 1, which revealed that older children tended to say "yes" to object questions but not to preference questions. One possible reason why the entities to which the questions pertained affected the children's yes bias might be the difficulty level of the questions. Both questions tested children's knowledge; however, responding to face questions, which tap social knowledge, may be easier than responding to object questions.

Six-year-olds already seemed to understand a variety of facial expressions and they were likely to insist on accurately expressing their opinions. For example, they sometimes said "No, he is smiling" to the card with happy face when asked "Is he glad?" and vice versa. This attention to specificity in their answers might explain why 6-year-olds tended to say "no" to face questions. They might understand that the experimenter was asking the questions to assess their knowledge, and they wanted to demonstrate a high level of knowledge. Thus, these children might understand the question interview context more deeply than the younger children..

Finally, the children in this experiment did not often respond with "I don't know" or with no response. One possibility is that we used object questions involving only familiar objects and face questions regarding highly familiar emotional expressions, and thus Japanese children were capable of answering these familiar questions (see also Okanda & Itakura, 2008). That the children were comfortable with the familiarity of these items is supported by the finding that the 5- and 6-year-olds in Study 2 did not display a yes bias to yes–no questions to familiar objects, in contrast to that found for children at this age in Study 1.

In this study, we have found that 3-year-olds exhibited a yes bias to both types of yes–no questions (object, face); however, the question that still remains unanswered is whether children tend to say "yes" to yes–no questions despite knowing the answer or because they do not know the answer. In

Study 3, we confirmed whether or not the 3-year-olds who participated in Study 2 had prior knowledge of the target entities of the questions.

Study 3

In this study, we investigated whether 3-year-old children knew the answer or did not know the answer to the yes–no questions when they responded with a "yes" response, and whether children's level of knowledge about the entity influenced their yes bias. The 3-year-old children were asked questions with two response options and were asked to indicate the correct answer.

Method

Participants

Participants of the study included twenty-two 3-year-old Japanese children (M = 40.6 months, SD = 3.39, range = 36–47 months; 11 boys and 11 girls) who had participated in Study 2.

Materials and Procedures

Each child was asked 18 choice questions relevant to objects and face stimuli from Study 1 and Study 2 and previous studies (Fritzley & Lee, 2003; Okanda & Itakura, 2007, 2008). The target objects included six familiar objects (i.e., blue cup, red apple, green apple, book, spoon, and toothbrush) and the eight face cards that were used in Study 2. Two objects or two cards were shown to the children, and they were instructed to indicate the correct answer from the two response options by pointing. The questions were based on the object condition or the face condition; the former had 10 questions and the latter had 8 questions.

Each child was tested individually. The conditions were counterbalanced and the order of questions was randomized. The experimenter chose one of the two conditions, object or face, and showed two objects or two face cards to the child. The experimenter then instructed the child to indicate one of the two options and asked the questions appropriate to the item (i.e., "Which is X?" or "Which is for X?"; Appendix C provides the list of questions asked in Study 3). The same two face cards were always presented together. The other condition followed the first condition using the same procedure. In order to avoid order effects, half of the children began Study 2 first, and the other half began Study 3 first. The one study immediately followed the other study.

Results

Each child received a score of 1 for a correct answer and a score of 0 for an incorrect answer. Subsequently, the proportional correct score of the choice questions was obtained by calculating the average of the scores, and thus, the overall scores of the responses to object and face choice questions were in the range 0.0–1.0. The children's face choice question mean scores were .98 (SD = .05), and object choice question mean scores were .97 (SD = .06).

The children's correct scores for the yes–no questions asked in Study 2 were also obtained in this study. The children received a score of 1 for a correct answer and score of 0 for an incorrect answer. Subsequently, the proportional score of the yes–no questions was obtained by calculating the average of the scores, and thus the overall scores of the responses to object and face yes–no questions were in the range 0.0–1.0. The children's face yes–no question scores were .43 (SD = .09) and object yes–no question scores were .38 (SD = .08).

A 2 × 2 within-design ANOVA (entity: object, face; question type: choice, yes–no) was conducted. The main effect of entity was significant, $F(1, 21) = 5.78$, $p < .05$. The children's performance was better when they were asked face questions than object questions. The main effect of question type was also significant, $F(1, 21) = 1367.77$, $p < .01$. The children gave significantly more correct answers to choice questions than to yes–no questions.

One-sample t tests were conducted to investigate whether the children were better able to choose the options or answer the yes–no questions correctly. The mean correct scores of the object and face conditions for the choice and yes–no questions were compared to a chance level (0.5). The children's mean scores for correct responses to the choice questions were significantly higher than the chance level: object condition, $t(21) = 35.14$, $p < .01$, and face condition, $t(21) = 46.81$, $p < .01$. The children were able to choose the correct answer from the two options. The children's mean scores for correct responses to the yes–no questions were significantly lower than the chance level: object condition, $t(21) = -7.37$, $p < .01$, and face condition, $t(21) = -3.69$, $p < .01$. Thus, the children were unable to correctly answer yes–no questions.

Discussion

Study 3 investigated whether children exhibited a yes bias even if they know the correct answer, and whether their level of knowledge about the entity influences their yes bias. Three-year-olds were able to provide correct answers when they were asked questions in the choice question format, and their knowledge about the entity seemed to influence their yes bias.

Our finding suggests that even though the children knew the answer, they tended to say "yes"; that is, they answered inappropriately only to yes–no questions. In addition, the children were found to be more knowledgeable about facial expressions than objects. The level of knowledge with regard to the entity could partly explain why children exhibited a weaker yes bias to questions about facial expressions, which they are more knowledgeable about, than they did to questions about objects.

General Discussion

The present study investigated whether 3- to 6-year-old Japanese preschoolers exhibit a yes bias to yes–no questions from two knowledge domains, and whether 3-year-olds possess sufficient knowledge about the subject of the questions. It is important to note that this was the first study, to our knowledge, that investigated whether children display a yes bias to questions about one's own preferential attitude and about the social entity of facial expressions. In Study 1, 3-year-olds exhibited a yes bias to both knowledge-object and preference-object questions, but 4-year-olds displayed a yes bias only to knowledge-object questions. Five- to 6-year-olds tended to exhibit a very weak yes bias that needs a further investigation. In Study 2, only 3-year-olds exhibited a yes bias to both knowledge-object and knowledge-face questions, and 6-year-olds showed a nay-saying bias to knowledge-face questions. In Study 3, 3-year-olds were able to answer choice questions that measured their knowledge of the objects, and level of knowledge about the entity might be found to influence their yes bias.

First, let us address the yes bias among 3-year-olds, who exhibited a strong yes bias to every yes–no questions even when they were knowledgeable about the entities to which the questions pertained. These results are partly consistent with earlier studies that revealed that 2-year-olds were unable to express what they know even when they have sufficient knowledge (e.g., Berthier, DeBlois, Poirier, Novak, & Clifton, 2000; Hood, Carey, & Prasada, 2000; Kloos, Haddad, & Keen, 2006). In these studies, 2-year-olds failed to find a ball on the ramp that was stopped by a barrier behind one of several opaque screens even though they seemed to have knowledge of solidity. Similarly, the children in this study might have failed to express their knowledge of objects and facial expressions and tended to inappropriately say "yes" to all questions.

Why do 3-year-olds exhibit a strong yes bias to yes–no questions? Understanding the context of questioning formats with adults and inhibitory control ability are considered plausible factors. Attention to the precise context of the question is needed to correctly answer the yes–no format of questions (Lee & Eskritt, 1999); however, 3-year-olds might not pay close attention to these contexts. Also, as Moriguchi et al. (2008) indicated, a "yes" response could be a dominant response because adults may not often ask children questions that require saying "no." Three-year-olds might not be able to inhibit their dominant "yes" response, and they automatically exhibited a yes bias (Moriguchi et al., 2008; Okanda & Itakura, 2007). However, it is also possible that the knowledge of 3-year-olds might not be as consolidated as that of older children. Choice questions probably required less knowledge than yes–no questions did because two options of the answer were presented in the choice questions but no such options in the yes–no questions. As the result, the children could perform well on the choice questions. Also, the issue of compliance is still a potential factor in accounting for children's yes bias. Children may be likely to say "yes" because adults are authoritative interviewers for them. In future research, we need to investigate whether children exhibit a yes bias to other children's yes–no questions. Children may not feel the same division of authority to children about the same age as they do with adults, and they may not show a yes bias to their peers.

Second, why do 4-year-olds and some older children exhibit a yes bias only to object questions? As discussed in Study 1, one possible explanation is that older children tend to display a yes bias to more difficult yes–no questions, and knowledge questions regarding objects appeared to be the most difficult type of the questions. Another possible explanation is that object questions which present a knowledge-testing situation may not be familiar to young children; hence, it might be difficult for the children to answer correctly. In fact, the children's yes bias was stronger when they were asked object questions than when they were asked face questions. Lee and Eskritt (1999) suggested that it is difficult for children to adjust their responses if they are asked questions in a context that violated daily conversations (see also Cordón, Saetermoe, & Goodman, 2003, 2005). It is likely that in day-to-day life, children come across questions regarding facial expressions and preferences more than they come across questions about objects. For example, mothers may often ask their children questions about a character's facial expression, or whether they like the character and contents, when they read a picture book. However, object yes–no questions used in the present and previous studies (Fritzley & Lee, 2003; Okanda & Itakura, 2007, 2008) asked children questions in an unusual manner. It may be rare for mothers to ask their children "Is this green?" while showing a red apple.

Third, we found that 6-year-olds exhibited a nay-saying bias to questions about facial expressions. They tended to insist on precisely expressing their opinions as accurately as possible and denied the interviewer's questions when asked face questions; however, they did not display the same tendencies to object and preference questions. A possible difference among these yes–no questions is that face questions might have several possible correct answers, but object and preference questions had very clear answers and were too easy for 6-year-olds to attempt to give more precise answers. Thus, one possible interpretation of this finding is that 6-year-olds might be able to understand the context of the question, that is, the interviewer wanted to know their knowledge, and they were likely to show the response tendency when asked questions that had several possible answers.

Fourth, object familiarity was not important for preference questions, but might be important for knowledge questions. In contrast, Okanda and Itakura (2007) found that object familiarity was not important for 2- to 3-year-olds when their own mothers asked them object yes–no questions. These results indicate that object familiarity may be important only when yes–no questions involving object knowledge are asked by a stranger.

Fifth, we did not find strong evidence that Japanese children tended to show "I don't know" and "no answer" responses to yes–no questions, similar to findings reported by Okanda and Itakura (2008). One possible reason might be that we mainly used familiar yes–no questions, and Japanese children are likely to avoid a binary yes–no decision when asked yes–no questions pertaining to unfamiliar issues (Okanda & Itakura, 2007, 2008). Moreover, our findings suggest there may be a cross-cultural difference with Japanese children's developmental pattern of a yes bias. While Fritzley and Lee (2003) reported that North American 4- and 5-year-olds did not show a yes bias, Japanese 4- to 6-year-old children sometimes showed a weak tendency to say "yes" (Okanda & Itakura, 2008; Study 1 in this article). We only found this tendency in Study 1, but not in Study 2, so it must be treated with caution; however, it is possible that Japanese children's yes bias disappears later than that of North American children. In Japanese culture, saying "yes" is much more polite than saying "no" in adult community. Possibly, some older children might have already adopted this Japanese attitude.

The present study provided new evidence for understanding the mechanism of a yes bias. Three-year-olds said "yes" to both types of yes–no questions, including testing-context yes–no questions (i.e., knowledge-object questions) and social-context yes–no questions (i.e., preference-object and knowledge-face questions). At the same time, we reported that 3-year-olds actually knew the answers to the questions when asked in the choice format and only failed to answer questions appropriately in the yes–no format. Underdeveloped cognitive abilities such as inhibitory control, ability to understand the context of the questions, and unconsolidated knowledge might be plausible explanations for a yes bias in 3-year-olds. In contrast, 4-year-olds and older children might have social reasons, such as sensitivity to interviewers' intentions and compliance, to show a yes bias. We conclude that there may be both cognitive and social reasons in the production of a yes bias, and its mechanism varies depending on the child's age.

On a practical level, and given that a yes–no question is the most common questioning method in developmental psychology (Fritzley & Lee, 2003), our results suggest that researchers should consider alternative methodologies in their experiments. Using yes–no questions for preschoolers is not suitable because they are likely to show different response tendencies to yes–no questions for different types of domains.

References

Berthier, N. E., DeBlois, S., Poirier, C. R., Novak, M. A., & Clifton, R. K. (2000). Where's the ball? Two- and three-year-olds reason about unseen events. *Developmental Psychology, 36*, 394–401.

Brady, M. S., Poole, D. A., Warren, A. R., & Jones, H. R. (1999). Young children's responses to yes-no questions:

Patterns and problems. *Applied Developmental Science, 3*, 47–57.

Cassel, W. S., Roebers, C. E. M., & Bjorklund, D. F. (1996). Developmental patterns of eyewitness responses to repeated and increasingly suggestive questions. *Journal of Experimental Child Psychology, 61*, 116–133.

Ceci, S. J., & Bruck, M. (1993). Suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin, 113*, 403–439.

Cordón, I. M., Saetermoe, C. L., & Goodman, G. S. (2003). Children in court. In S. Penrod & P. J. van Koppen (Eds.), *Adversarial versus inquisitorial justice: Psychological perspectives on criminal justice systems* (pp. 167–189). New York: Plenum.

Cordón, I. M., Saetermoe, C. L., & Goodman, G. S. (2005). Facilitating children's accurate responses: Conversational rules and interview style. *Applied Cognitive Psychology, 19*, 249–266.

Flavell, J. H., Flavell, E. R., & Green, F. L. (1983). Development of the appearance-reality distinction. *Cognitive Psychology, 15*, 95–120.

Flavell, J. H., Green, F. L., & Flavell, E. R. (1986). Development of knowledge about the appearance-reality distinction. *Monographs of the Society for Research in Child Development, 51*, 1–69.

Flavell, J. H., Zhang, X. D., Zoui, H., Dong, Q., & Qi, S. (1983). A comparison between the development of the appearance-reality distinction in the People's Republic of China and the United States. *Cognitive Psychology, 15*, 459–466.

Fritzley, V. H., & Lee, K. (2003). Do young children always say yes to yes-no question? A metadevelopmental study of the affirmation bias. *Child Development, 74*, 1297–1313.

Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology, 37*, 620–629.

Hood, B., Carey, S., & Prasada, S. (2000). Predicting the outcomes of physical events: Two-year-olds fail to reveal knowledge of solidity and support. *Child Development, 71*, 1540–1554.

Hughes, M., & Grieve, R. (1980). On asking children bizarre questions. *First Language, 1*, 149–160.

Kloos, H., Haddad, J. M., & Keen, R. (2006). Which cues are available to 24-month-olds? Evidence from point-of gaze measures during search *Infant Behavior & Development, 29*, 243–250.

Krähenbühl, S., & Blades, M. (2006). The effect of interviewing techniques on young children's responses to questions. *Child: Care, Health and Development, 32*, 321–331.

Lee, K., & Eskritt, M. (1999). Beyond the Gricean maxims: Conversational awareness as a multifaceted domain of knowledge. *Developmental Science, 2*, 27–28.

Moriguchi, Y., Okanda, M., & Itakura, S. (2008). Individual differences in children's yes bias and the relationship to verbal ability and inhibitory control. *First Language, 28*, 431–442.

Moston, A. (1990). How children interpret and respond to questions: Situational sources of suggestibility in eyewitness interviews. *Social Behaviour, 5*, 155–167.

Okanda, M., & Itakura, S. (2007). Do Japanese children say "yes" to their mothers? A naturalistic study of response bias in parent–toddler conversations *First Language, 27*, 421–429.

Okanda, M., & Itakura, S. (2008). Children in Asian cultures say yes to yes-no questions: Common and cultural differences between Vietnamese and Japanese children. *International Journal of Behavioral Development, 32*, 131–136.

Peterson, C., & Biggs, M. (1997). Interviewing children about trauma: Problem with "specific" questions. *Journal of Traumatic Stress, 10*, 279–290.

Peterson, C., Dowden, C., & Tobin, J. (1999). Interviewing preschoolers: Comparisons of yes/no and wh- questions. *Law and Human Behavior, 23*, 539–555.

Peterson, C., & Grant, M. (2001). Forced-choice: Are forensic interviewers asking the right questions? *Canadian Journal of Behavioral Science, 33*, 118–127.

Siegal, M., & Surian, L. (2007). Conversational understanding in young children. In E. Hoff & M. Shatz (Eds.), *Blackwell handbook of language development* (pp. 304–323). Oxford, UK: Blackwell.

Steffensen, M. S. (1978). Satisfying inquisitive adults: Some simple methods of answering yes-no questions. *Journal of Child Language, 5*, 221–236.

Waterman, A. H., Blades, M., & Spencer, C. (2001). Interviewing children and adults: The effect of question format on the tendency to speculate. *Applied Cognitive Psychology, 15*, 521–531.

Waterman, A. H., Blades, M., & Spencer, C. (2004). Indicating when you do not know the answer: The effect of question format and interviewer knowledge on children's "don't know" responses. *British Journal of Developmental Psychology, 22*, 335–348.

Zajac, R., Gross, J., & Hayne, H. (2003). Asked and answered: Questioning children in the courtroom. *Psychiatry, Psychology and Law, 10*, 199–209.

Zajac, R., & Hayne, H. (2003). I don't think that's what really happened: The effect of cross-examination on the accuracy of children's reports. *Journal of Experimental Psychology: Applied, 9*, 187–195.

Appendix A: Objects Used and Test Questions in Study 1

Familiar object condition		Unfamiliar object condition	
Object	Questions	Object	Question
Blue cup	Is this blue?	Coffee filter (plastic)	Is this used for making coffee?
	Is this used for drinking?		Is this empty?
	Is this made of glass?		Is this used for making cake?
	Is there water in this?		Is this made of paper?
	Do you like this?		Do you like this?
	Do you dislike this?		Do you dislike this?
Red apple	Is this hard?	Shoehorn	Is this used for wearing shoes?
	Is this for eating?		Is this found at the entrance?
	Is this rotten?		Is this worn on the head?
	Is this green?		Is this soft?
	Do you like this?		Do you like this?
	Do you dislike this?		Do you dislike this?
Book	Is this full of pictures?	CPU	Is this square-shaped?
	Is this used for reading?		Is this used with a computer?
	Is this small?		Is this made of wood?
	Is this round?		Is this circular?
	Do you like this?		Do you like this?
	Do you dislike this?		Do you dislike this?

Appendix B: Objects and Cards Used and Test Questions Asked in Study 2

Face condition		Object condition	
Target face	Questions	Target object	Questions
Face 1 smiling	Is it joyful?	Blue cup	Is this blue?
	Is it sad?		Is this used for drinking?
Face 1 crying	Is it annoyed?		Is this made of glass?
	Is it pleased?		Is there water in this?
Face 2 smiling	Is it happy?	Red apple	Is this hard?
	Is it lonely?		Is this for eating?
Face 2 crying	Is it crying?		Is this rotten?
	Is it smiling?		Is this green?
Face 3 smiling	Is it joyful?	Picture book	Is this full of pictures?
	Is it sad?		Is this used for reading?
Face 3 crying	Is it unpleasant?		Is this small?
	Is it joyful?		Is this round?
Face 4 smiling	Is it happy?	Spoon	Is this blue?
	Is it lonely?		Is this used for eating?
Face 4 crying	Is it crying?		Is this made of paper?
	Is it smiling?		Is this square?

Appendix C: Objects and Cards Used and Test Questions Asked in Study 3

Condition	Object/cards	Question
Familiar object	Blue cup/red apple	Which is used for drinking? Which is blue?
	Red apple/green apple	Which is red? Which is green?
	Picture book/red apple	Which is used for reading? Which is for eating?
	Spoon/picture book	Which is used for eating? Which has pictures?
	Toothbrush/spoon	Which is used for brushing teeth? Which is placed in the bathroom?
Face	Face 1 happy/sad	Which is joyful? Which is displeased?
	Face 2 happy/sad	Which is pleased? Which is sad?
	Face 3 happy/sad	Which is crying? Which is smiling?
	Face 4 happy/sad	Which is happy? Which is lonely?

Child Development, March/April 2010, Volume 81, Number 2, Pages 581–596

Lying and Truth-Telling in Children: From Concept to Action

Fen Xu
BNU State Key Laboratory of Cognitive Neuroscience and Learning

Xuehua Bao
Zhejiang University

Genyue Fu
Zhejiang Normal University

Victoria Talwar
McGill University

Kang Lee
University of Toronto and University of California, San Diego

Although there has been extensive research on children's moral knowledge about lying and truth-telling and their actual lie- or truth-telling behaviors, research to examine the relation between these two is extremely rare. This study examined one hundred and twenty 7-, 9-, and 11-year-olds' moral understanding of lies and their actual lying behaviors in a politeness situation. Results revealed that as age increased, children increasingly evaluated others' lying in politeness situations less negatively and were more inclined to tell lies in such situations themselves. Contrary to previous findings, children's sociomoral knowledge about lying was significantly related to their actual behaviors, particularly when children's rationales underlying their moral judgments were consistent with their motives for actual lie- or truth-telling in the politeness situation.

Lying is a common social phenomenon. It occurs regularly in various social contexts for a multitude of purposes. For children, there are two types of lies that are of great importance during their socialization. One type is the lies that violate moral rules as they are typically told to benefit oneself at the expense of others. As a result of the antisocial nature of this type of lie, it is universally discouraged by children's caregivers and teachers from a very early age. The other type of lies are those that are told with an intention to help, not harm, another individual (e.g., faking liking an undesirable gift in front of a gift-giver) and are thus prosocial in nature. Although philosophers and theologians have long debated about whether prosocial lies should be morally sanctioned (Bok, 1978), in everyday practice, such lies are told frequently (DePaulo & Bell, 1996; DePaulo & Kashy, 1998) and often entail positive values (Nyberg, 1993). Some theorists such as Sweetser (1987) have even suggested that in some situations prosocial lies are not only socially acceptable but also are not lies at all.

There has been extensive research on the development of lying that dates back to the beginning of developmental psychology (e.g., Binet, 1896; Darwin, 1877; Piaget, 1932). This long-standing interest in this topic is a result of the fact that lying can serve as a window into many aspects of children's developing minds, for example, intelligence (Binet, 1896; Lewis, 1993), theory of mind (Chandler, Fritz, & Hala, 1989; Leekam, 1993; Peskin, 1992; Peterson, 1995; Peterson, Peterson, & Seeto, 1983; Polak & Harris, 1999; Sodian, 1991; Talwar, Gordon, & Lee, 2007), moral understanding (Piaget, 1932), personality and character formation (Hartshorne & May, 1928), and children's competence as witnesses in the courts of law (Goodman et al., 2006; Lyon, 2000; Strichartz & Burton, 1990; Talwar, Lee, Bala, & Lindsay, 2002, 2004; for a review, see Lee, 2000).

Research on Antisocial Lying

To date, most studies have focused on the development of antisocial lies and can be divided into

---

This study was supported by grants from NIH (R01 HD048962), the University of California Pacific Rim Program, and the Social Sciences and Humanities Council of Canada to K. Lee.

Correspondence concerning this article should be addressed to Kang Lee, Institute of Child Study, 45 Walmer Road, Toronto, Ontario, Canada M5R 2X2. Electronic mail may be sent to kang.lee@utoronto.ca.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8102-0014

two major streams. One stream investigates children's conceptions and moral judgments of antisocial lies. Extensive research has revealed that children from very early on are taught explicitly or implicitly about the negative moral implications associated with antisocial lies. They are also strongly discouraged from telling this type of lie (Cameron, Chen, & Lee, 2001; Wilson, Smith, & Ross, 2003). Perhaps for this reason, the conceptual understanding of antisocial type of lies emerges as early as 3 years of age (Siegal & Peterson, 1998; Talwar et al., 2002). With regard to children's moral judgments of lying, evidence shows that even preschoolers understand that antisocial lies entail negative moral implications (Bussey, 1992, 1999; Talwar et al., 2002).

The other stream of research, pioneered by Binet (1896), Darwin (1877), and Hartshorne and May (1928), examines children's lying behavior for trickery, personal protection, or benefit (Chandler et al., 1989; Lewis, Stanger, & Sullivan, 1989; Peskin, 1992; Polak & Harris, 1999; Talwar & Lee, 2002a). It has been found that even preschoolers tell lies with intent to deceive (Chandler et al., 1989; Peskin, 1992). Young child lie-tellers are generally skilled in masking their nonverbal behaviors (e.g., making eye contact while lying; Lewis et al., 1989; Talwar & Lee, 2002a) but poor at concealing their deception in their verbal statements (e.g., blurting out the name of a toy that they claimed not to have peeked at: Polak & Harris, 1999; Talwar & Lee, 2002a, 2008; Talwar, Gordon, et al., 2007).

Research on Prosocial Lying

In contrast to the extensive literature on antisocial lying, there is very limited research on the development of lie-telling for politeness purposes. Research on this issue should offer a unique perspective for understanding social development in children owing to the multifaceted nature of prosocial lie-telling. More specifically, to take an appropriate action in a politeness situation, one must assess, and reason about, both moral and social conventional rules regarding interpersonal interaction. Turiel and his colleagues (Helwig & Turiel, 2002; Nucci & Turiel, 2000; Smetana & Turiel, 2003; Turiel, 2002) have found that moral and social conventional rules are distinctive domains in the realm of social life and social thought. Moral rules are concerned with what is right and what is wrong as well as regulating how individuals ought to behave toward each other based on the principles of welfare and justice. In contrast, social conventional rules are concerned with uniformities that define social expectations regarding appropriate behaviors within a specific society and serve functions of social coordination. Typically, these rules deal with individuals' social thoughts and actions in their respective domains. For example, moral rules prohibit killing of another human being, whereas social conventional rules may require one to dress properly at a formal social gathering.

Politeness, however, cuts across both domains. In the politeness situation, one must assess, and reason about, the applicability of various moral and social conventional rules to take a morally and socially appropriate action. More significantly, the politeness situation brings to the fore the inherent contradictions in the rule system in either the moral or social conventional domains, respectively. In the moral domain, the concern for others' well-being and the need to avoid harm should motivate one to tell a prosocial lie (e.g., "I really like your gift" or "You look great in that dress") rather than telling the blunt truth (e.g., "I do not like your gift" or "You look fat in that dress"). However, lying, or making a false statement with an intent to deceive, has long been considered to be a serious moral transgression by philosophers and theologians, such as St. Augustine (1952), Kant (1949), and Bok (1978), because lying impinges on the lie recipient's right to information and freedom of choice (Grice, 1980; Krupfer, 1982). Children have also been socialized to take this view of lying from very early on in life (Cameron et al., 2001; Wilson et al., 2003).

In the social-conventional domain, there are also contradictory rules concerning interpersonal communication, particularly in the case of politeness are clearly demonstrated. On one hand, Grice (1980) suggested that one of the most fundamental conventions governing interpersonal communication is the Maxim of Quality. This maxim requires speakers to inform, not misinform, their communicative partners, which is commonly assumed to be the case when individuals enter into a conversation. Prosocial lies clearly flout this maxim. On the other hand, Lakoff (1973) proposed that interpersonal communications must also adhere to the Rule of Politeness that requires speakers to be amicable to their communicative partners. In the politeness situation, this rule calls for individuals to not tell the blunt truth to establish or maintain an amicable relationship with another.

The existing evidence suggests that many adults have little difficulty in resolving the conflicts evoked by the politeness situation. Typically, they allow that the need to be polite and to avoid hurting others overrides the need to be truthful. For this reason, adults not only endorse prosocial lies (Lee & Ross, 1997) but tell them regularly (DePaulo & Bell, 1996; DePaulo & Kashy, 1998). In contrast, for children, the politeness situation poses a challenge to children who are still in the process of acquiring moral and social-conventional rules of their society. Resolving conflicts between these rules places additional demands on them. Although the politeness situation may be a challenge for developing children, it offers a naturalistic opportunity for developmental researchers to understand whether children are capable of making strategic trade-offs when facing social situations where rules are in conflict as well as how they select and apply rules adaptively in different social situations.

To date, there are only three studies that have examined children's actual behavior in a politeness situation. Talwar and Lee (2002b) used a Reverse Rouge task in which the experimenter had a conspicuous mark of lipstick on the nose. The child was asked to take a picture of the experimenter, but before the picture was taken, the experimenter asked, "Do I look okay for the picture?" Results showed that 89% of children between 3 and 7 years of age stated that the experimenter looked okay. However, when the experimenter left, children told another adult that the experimenter actually did not look okay. Thus, the researchers concluded that young children can tell prosocial lies in a politeness situation. However, because children were not probed about why they told such a lie, it is unclear whether the children in the study considered the contradictory rules evoked by the situation when deciding to lie. Furthermore, among the children who lied, it was unclear as to whether they told lies to spare the feelings of the experimenter (a prosocial lie) or to avoid potential negative consequences if the truth was told (a self-protective lie).

In another study (Fu & Lee, 2007), Chinese children aged between 3 and 6 years were asked to rate pictures that were poorly drawn by confederates. Most children, except for the youngest ones, gave the drawings more positive ratings in front of the confederates than when the confederates were absent. Thus, like Western children, Chinese preschoolers appeared also to refrain from telling the blunt truth to others (i.e., the pictures were poorly drawn). However, because the children were not probed about their motivation for inflating their ratings, it is not clear whether they did so for the benefit of the confederate (e.g., sparing the feelings of the confederate) or for themselves (e.g., avoiding negative consequences if the truth was told).

In the third study, Talwar, Murphy, and Lee (2007) used a disappointing gift paradigm to examine children's prosocial lie-telling. Children played a game where they were promised a gift from a gift basket that contained a range of different toys and gifts. After the task, children received an undesirable gift of soap instead of a toy and were questioned by the gift-giver about whether they liked the gift. In this situation, children had to reconcile their desire for a better gift with the competing social and moral requirement to be polite. When asked if they liked the gift, the majority of children told the gift-giver untruthfully that they liked the disappointing gift, despite having told their parents that they did not like the gift. School-aged children were more likely to lie than were preschool children. However, this study also did not systematically probe children's justifications for their lies. Thus, it is again not clear whether some children were more motivated to lie for self-protection (e.g., avoiding negative responses from the gift-giver if the truth was told) or to be polite and protect the gift-giver's feelings.

In contrast to the paucity of research on children's actual prosocial lie-telling behaviors, there is some, albeit limited, research on children's conceptual understanding of prosocial lies. With regard to children's concept of prosocial lies, Lee and Ross (1997) found that adolescents aged between 12 and 14 years and college students were less inclined to classify untruthful statements as lies when told with the intent to help another individual than when told to harm. Their results confirmed Sweetser's (1987) theoretical contention that untruthful statements deliberately told to help another individual and to be polite may not be considered lies. In contrast, Bussey (1999) reported that most children aged between 4 and 11 years classify all types of untruthful statements as lies regardless of their anti- or prosocial nature. This finding suggests a possible developmental change in terms of the concept of prosocial lies during adolescence.

With regard to moral judgments, Bussey (1999) found that children during preschool and elementary school years tended to give negative ratings to prosocial lies. Nevertheless, beginning from 4 years of age, their ratings of prosocial lies were not as negative as those given to trickery or antisocial lies. Broomfield, Robinson, and Robinson (2002) further found that children aged between 4 and 9 years would suggest that a story character should tell a lie about liking an unwanted gift to make the giver believe the gift was liked. The children also judged that the gift-giver would be happy when hearing the lie. The results of Broomfield et al. (2002) were replicated in China (Zeng, 2004). However, Walper and Valtin (1992) found that children only began to give prosocial lies positive evaluations at the end of the elementary school years. These findings taken together suggest that children's conceptual understanding of prosocial lies begins in preschool years and develops throughout the childhood. They appear to be able to consider the contradictory rules evoked by the politeness situation when evaluating prosocial lies, but only in late childhood do children appear to allow the need to be polite and avoid hurting another's feelings override the need to be truthful. It should be noted that this conclusion is tentative because these studies did not probe children about the underlying rationales for their evaluations.

No studies have examined the relation between children's conceptions of prosocial lie-telling and their actual lying behavior. Research on the general relation between children's social and moral conceptions and their actual behaviors is of significant importance because the ultimate purpose of socialization is to ensure that children not only know morally what is right or wrong and conventionally what is appropriate or inappropriate, but that they also act accordingly. Existing studies with children and adults concerning Kolhbergian moral dilemmas as well as antisocial lying have shown either weak or no linkages between individuals' moral knowledge and moral action (Arnold, 1989; Blasi, 1980; Talwar et al., 2002, 2004; Thoma & Rest, 1986). Multiple theories have provided a variety of explanations as to why individuals fail to act according to their conceptual knowledge about moral or social rules (Arsenio & Lemerise, 2004; Crick & Dodge, 1994; Huesmann, 1998; Rest, 1986; Walker, 2002), including situations involving aggression (Crick & Dodge, 1994; Huesmann, 1998) or prosocial behaviors (e.g., Rest, 1986). A major factor contributing to this disconnection is the difference in the interpretation of hypothetical situations used to assess moral and social *knowledge* and interpretations of actual situations that call for morally and socially appropriate *behavior*. This is particularly true for situations where contradictory moral and social rules may apply. For example, in a politeness situation, individuals may interpret the need to be polite and to avoid hurting another to be the primary concern and thus advocate hypothetically that one should tell a prosocial lie in such a situation. However, when encountering the same situation in reality, they may interpret the situation to require only for the fulfillment of the need to be truthful and thus tell the truth. Alternatively, individuals could interpret the hypothetical situation to be concerned with the need to the truthful but the real-life situation to be concerned with the need to be polite and to avoid hurting another. In both cases, a disconnection between conceptual knowledge and action results.

Thus, to examine empirically the relation between children's *conceptual knowledge* about prosocial lie-telling and their *actual* lying behavior, two methodological measures must be taken. First, one must examine children's interpretations of hypothetical politeness stories that are used to assess their understanding of prosocial lie-telling. Second, one must also obtain children's interpretations of their own actions in the politeness situation. Such methodological measures were not taken in the previous studies, which leaves gaps in our knowledge about: (a) the age at which children begin to understand the politeness situation to be a situation where one needs to consider the need to be truthful and the need to be polite and avoid hurting another, (b) the age at which they begin to tell lies for prosocial reasons, and most importantly, (c) whether children's conceptual understanding of prosocial lie-telling is related to their actual behaviors in the politeness situation. This study was conducted to bridge the gap in the literature and to address these three important questions.

In the present study, we assessed 7-, 9-, and 11-year-old Chinese children's conceptual understanding of prosocial lie-telling and blunt truth-telling in a politeness situation. The three age groups were chosen because existing studies have shown that children's conceptual knowledge about prosocial lie-telling appears to undergo systematic changes during this period (e.g., Broomfield et al., 2002; Bussey, 1999; Walper & Valtin, 1992; Zeng, 2004). Also, this developmental trend appears to be similar between Western and Chinese children (see Broomfield et al., 2002; Zeng, 2004). The Chinese children were read stories in which story characters encounter politeness situations (e.g., receiving an undesirable gift). The story characters either made truthful or untruthful statements. Children were asked to classify the statements as lies or truths. Also, they were asked to evaluate the statements either positively or negatively and to justify their ratings. Children were also asked to respond to control stories in which a story character receives a desirable gift and likes it and tells the truth or lie about it. The use of these control stories was to ensure that children's categorizations and moral judgments of lie- or truth-telling about undesirable gifts would not be simply responses to truthful and untruthful statements in general.

Furthermore, we placed the children in a real-life situation where they themselves must decide whether to tell the truth or lie. In this situation, children were given an undesirable gift and then asked by the gift-giver whether they liked the gift. This method was a modified version of the undesirable gift paradigm pioneered by Saarni (1984) and Cole (1986) to examine children's expressive display rule use, and adapted for examining children's lie-telling by Talwar, Murphy, et al. (2007). The undesirable gift situation was chosen because this is a situation that occurs commonly in children's lives (e.g., receiving a poorly fitting knitted sweater from grandparents as a birthday gift), and children are socialized from early on to dissemble their true feeling of disliking in such situations. Indeed, from 4 years of age, children are already able to dissemble their nonverbal (Cole, 1986; Saarni, 1984) or verbal behaviors (Talwar, Murphy, et al., 2007) to appear pleased about receiving an undesirable gift.

Given the previous findings, we expected children in our study would also dissemble verbally and tell the gift-giver a prosocial lie that they liked the undesirable gift. This tendency to lie would increase with age. With regard to children's concept of prosocial lie-telling, based on the results of Bussey (1999) and Lee and Ross (1997), we predicted that children in general would classify untruthful statements told for prosocial purposes as lies and this tendency would decrease with age. With regard to children's evaluations of prosocial lie-telling, based on the results of Bussey (1999) and Walper and Valtin (1992), as age increased, children would be more inclined to rate prosocial lies less negatively than lies told for nonprosocial reasons. Also, with increased age, children would increasingly weigh the relative importance of being truthful versus being polite and avoiding hurting another when making evaluations and use the latter rather than the former to justify their ratings. Finally and most importantly, a significant relation between children's conceptions of prosocial lies and their actual behavior was expected. We hypothesized that this possible relation can be established when children's motives for their actions are consistent with the rationales underlying their moral judgments. For this reason, after children in this study told either the truth or a lie, they were asked about the reasons for their actions. We expected that the relation between children's conceptual understanding of prosocial lie-telling and their actual behavior would be stronger when the concerns for politeness and avoidance of hurting others motivated both their evaluations and their actions.

Method

Participants

A total of 120 Chinese children participated: forty 7-year-olds (M = 7.30 years, SD = 0.19; 20 boys), forty 9-year-olds (M = 9.18 years, SD = 0.14; 20 boys), and forty 11-year-olds (M = 11.21 years, SD = 0.13; 20 boys). The children were recruited from a large elementary school of close to 1,000 students in a southeastern city in P.R. China (population: 10 million). The school was chosen because it was highly representative of an elementary school in the city in terms of its ethnic composition as well as parental income, occupation, and educational levels according to the school records. Student admission to the school was determined by the location of the neighborhood where the children and their family lived rather than other criteria (e.g., parental professional standing, job affiliation, or the child's scores of entrance examination). The neighborhood from which the children came was highly diverse in terms of parental income, occupation, and educational levels. The children were Han Chinese, which represents 98% of the population in the city and 90% of the P.R. China population. We did not measure the socioeconomic status of the participants' family because no reliable and accurate measures were available. The children were selected from their classes based on the following factors: guardian consent, home classroom teacher's consent, the children's own verbal assent, and the requirement to balance the numbers for boys and girls per age group.

Materials and Procedures

Children were seen individually in a quiet room in their school. They participated in two sessions: One session examined whether children would tell a prosocial lie (undesirable gift procedure), and the other assessed children's conceptual understanding of lies and truths in politeness situations (moral story procedure). The two sessions were separated by 10 days and counterbalanced between subjects. Half of the children participated in the undesirable gift procedure first followed by the moral story procedure, and the other half did the opposite.

Moral story procedure. Experimenter 1 introduced herself as a student teacher from a local university and read children four stories accompanied by pictures. Before beginning the stories, children were first trained about how to use a 7-point Likert scale. The positive, neutral, and negative ratings were symbolized by red stars, a circle, and black crosses as follows: very very good (three stars), very good (two stars), good (one star), neither good nor bad (a circle), bad (one cross), very bad (two crosses), and very very bad (three crosses). Children's verbal responses were then converted into a 7-point numeric scale from −3 to 3 with 3 representing *very very good*, 0 representing *neither good nor bad*, and −3 representing *very very bad*.

In the stories, a child protagonist interacted with a teacher from the school. The stories were modified from those used in the previous studies on prosocial lie-telling (Bussey, 1992; Lee & Ross, 1997; Walper & Valtin, 1992). In one story, a child receives a gift that he dislikes but tells his teacher, the gift-giver, that he likes the gift (the prosocial lie-telling story). In the second story, a child receives an undesirable gift and tells the teacher the truth (the blunt truth-telling story). In the third story, a child receives a desirable gift but states falsely that she does not like the gift (the control lie-telling story). In the fourth story, a child receives a desirable gift and states truthfully that she likes the gift (the control truth-telling story). Existing studies (e.g., Bussey, 1992; Talwar et al., 2002) have consistently shown that children from preschool years tend to rate lying depicted in the control lie-telling story very negatively and truth-telling in the control truth-telling story very positively. Children's ratings for these two stories thus could be used to be compared to their ratings for the prosocial lie-telling and blunt truth-telling stories, respectively. The presentation of the stories was counterbalanced across subjects. The moral story procedure lasted about 10 min.

The actual stories are described in the following.

1. *Prosocial lie-telling story*: This is Ming Ming. He did very well in a math competition. His teacher gave him a pencil box as a gift. But Ming Ming already had many pencil boxes at home. He did not want any more of this kind of pencil box. So Ming Ming didn't like the gift his teacher gave him. The teacher asked him: "Do you like the pencil box?" Ming Ming said: "Yes, I like it very much."
2. *Blunt truth-telling story*: The final exam was finished and Xiao Qiang did very well. The teacher gave a book to Xiao Qiang as a gift when she distributed the grade report. When Xiao Qiang received the book, she found she already had this book at home. She already knew the book very well and wanted to read a new book. So Xiao Qiang didn't like the gift that the teacher gave to her. The teacher asked Xiao Qiang, "Do you like the book?" Xiao Qiang answered, "No, I don't like it."
3. *Control lie-telling story*: The teacher knew that many students in the class would celebrate their birthdays. Today is Xiao Hao's birthday. So she gave Xiao Hao an eraser as a birthday gift. Xiao Hao received the eraser and thought it looked cute. He liked the eraser. The teacher asked Xiao Hao, "Do you like the gift?" Xiao Hao said to the teacher, "No, I don't like the eraser."
4. *Control truth-telling story*: The teacher brought some landscape color pictures from home as a gift to give to the students in the class. Li Li thought these pictures were beautiful. Li Li was a representative of the class and went to the office to turn in the students' homework after class. The teacher wanted to know if the students liked the pictures. She had asked many of her students, but hadn't asked Li Li. So the teacher stopped Li Li and asked, "Do you like the picture I give you?" Li Li said, "Yes, I like it."

After being told each of the stories, the participants were asked the following questions: (a) Question 1 (categorization question): "Is what XXX (the name of the protagonist) said a lie or not a lie?" (If the participant replied "not a lie," they were asked "Is what XXX said the truth or not truth?"); (b) Question 2 (moral evaluation question): "Is what XXX said very very good, very good, good, neither good nor not bad, bad, very bad, or very very bad?" Children were asked to indicate their responses on the Likert scale they had been trained on. Finally, children were asked to justify their ratings (rating justification question): "Why do you think it is XXX (the child's rating)?"

The coding system of children's justifications was developed in the following manner. First, the potential types of responses were obtained based on the reports and discussions of the several existing studies (Bussey, 1999; Peterson et al., 1983; Walper & Valtin, 1992). Second, all types of responses provided by participants were reviewed by the research team members to obtain additional types of responses if any existed. Third, because participants did not provide a diverse array of justifications, the final coding system only contained the following types of response categories. The first type referred the need to be polite and protect another person's feelings as justification (politeness justification; e.g., ''It is not polite if I say I don't like it'' or ''It would not be nice'' or ''It would be embarrassing''). The second type made reference to the need to be truthful (honesty justification; e.g., ''He didn't like it'' or ''You should tell the truth''). The third type made reference to both politeness and honesty justifications (both justifications; e.g., ''Saying liking it will make the teacher feel better, but it is lying''). We also created an ''other'' category for responses (e.g., ''I don't know'') that did not fall into one of the three categories. Two graduate students who were naïve to the hypotheses of the study were trained to code children's responses according to these categories.

Undesirable gift procedure. In the procedure, children were seen individually by Experimenter 2 in a quiet room in their school. The child's teacher introduced the experimenter as a student teacher from a local university. The experimenter brought children to a quiet room in their school. She told the participants that they were going to take two memory tests. If they did well on both tests, they would receive a gift selected from a gift box. Children were shown each item from the box. They were ball-point pens, stickers, coloring books, flash cards, and pencils. Pilot testing showed that school-aged children liked the attractive and colorful pens, stickers, and coloring books, but disliked the plain-looking flash cards and pencils. Children were asked which item they would like to receive as a gift and which they would not like to receive (the Gift Preference Assessment). Consistent with the findings of the pilot study, no child liked the flash cards and pencils. Upon identifying the liked and disliked gift items, children were told that they would receive the gift item they liked the most if they passed the tests. Then, the experimenter administered a short-term memory test (a subtizing task) in which a number of pencils were shown briefly and children were asked to report how many pencils they saw. To ensure all children succeed in the test, the number of pencils shown was smaller than what would be expected to be remembered by children at the same age as the participants, which was determined by pilot testing. After all participants passed the test, they were praised. The experimenter told the children that she just remembered that she had to finish unspecified business and had to leave the room for a few minutes. Their own teacher would have to administer the second test and give them the gift if they also passed the second test.

After the participants' own teacher entered the room, the experimenter left. The teacher administered a digit span test similar to the subscale in Wechsler Intelligence Scale for Children, Fourth Edition (WISC). Based on the norms of the WISC, only those test items that were known to be easy for the participants' age were chosen. After all the participants passed the second test, the teacher said to the children: ''You did very well on both tests, so I am going to give you a gift.'' She pointed to the gift items in the gift box and said: ''The student teacher bought some of the gifts (pointing to the attractive ones). I bought these gifts (pointing to the unattractive ones). Let me give you the one I bought as a gift.'' Then, she handed over an unattractive gift to the child. After the child had the opportunity to inspect the gift, the teacher asked: ''Do you like the gift I have given to you?'' (Liking Question 1).

After the children received the gift and responded to the Liking Question 1, the experimenter returned. The teacher told the experimenter that the children did very well on the second test and therefore received a gift (pointing to the gift). She also informed the experimenter whether the children said that they liked the gift or they did not like the gift. The teacher then made an excuse and left the room. At this point, if children said that they liked the gift, the experimenter asked the children: ''I just heard that your teacher said that you told her you liked this gift but you told me earlier that you didn't like the gift. So, do you really like the gift or not?'' (Liking Question 2). If the children responded that they did not like the gift, the experimenter asked: ''Why did you tell your teacher you liked the gift?'' (action justification question). If the children responded directly to their teacher that they did not like the gift, the teacher also informed the experimenter of the children's response. After the teacher left, the experimenter also asked the children to justify their response: ''Why did you tell your teacher you didn't like the gift?'' After this question, the children were told that they could exchange their gift if they wished to do so. All children opted to do so. They were thanked for their participation in the study. The undesirable gift procedure lasted about 10 min. All experimenters and teachers were women. They were trained about how to carry out the procedures prior to interviewing children. The experimenters were randomly assigned to be Experimenter 1 or 2.

The children's responses to the "why" question were coded according to a coding system developed ad hoc based on the responses because no previous studies had coded such responses. Once the research team members reviewed the responses, it became clear that although children who told the truth consistently made reference to the need to be truthful or made a statement of the fact (e.g., "You should tell the truth" or "I don't like it"), among children who told a lie to conceal their disliking of the undesirable gift, some were not motivated to lie out of prosocial concerns. Thus, the following categories emerged for the children who lied: (a) the prosocial lie-tellers who justified their lying by referring to the need to be polite or to avoid hurting the feelings of their teacher, and (b) nonprosocial lie-tellers who gave a variety of responses unrelated to prosocial concerns such as referring to the need to avoid negative consequences to themselves (e.g., the teacher being angry at them if truth be told) or deserving the gift even though they did not like it, or other miscellaneous responses unrelated to politeness (e.g., "I don't know" or "I can find another use for it"). The same two graduate students were trained to categorize children into three categories: blunt truth-tellers, prosocial lie-tellers, and nonprosocial lie-tellers based on whether the children told the truth or lie and their responses to the "why" question.

Results

In the following, we will first report results about children's knowledge of lie- and truth-telling as assessed by the moral story procedure, followed by results about children's actual lie- or truth-telling behavior in the undesirable gift procedure, and finally results about the relations between children's knowledge and behavior. Preliminary analyses revealed that the order of the moral story procedure and undesirable gift procedure and the sex of participants did not significantly affect the results of this study. Thus, the data for these two factors were combined for the subsequent analyses.

Moral Story Procedure

Children's classifications of the statements made by the story characters as lies or truth or neither (i.e., their responses to the categorization question) were analyzed using multiple chi-square analyses. Neither the age group nor the story type effect was significant. Overall, 98% of the children categorized prosocial lies as lies, and 100% categorized the untruthful statements in the control lie-telling story as lies. Also, 98% categorized the blunt truths as the truth, and 100% categorized the truthful statement in the control truth-telling story as the truth.

Figure 1 shows the means and standard errors of 7-, 9-, and 11-year-olds' moral evaluations of the truthful or untruthful statements made by the story characters (i.e., their responses to the moral evaluation question). An omnibus 2 (truthfulness: truth vs. lie) × 2 (setting: prosocial vs. control) × 3 (age: 7, 9, and 11 years) analysis of variance (ANOVA) was performed on the moral evaluation data for the truth- and lie-telling stories.

Age, truthfulness, and setting main effects were significant, $F(2, 117) = 3.12$, $p < .05$, $\eta^2 = .05$; $F(1, 117) = 264.82$, $p < .0001$, $\eta^2 = .69$; and $F(2, 117) = 14.25$, $p < .0001$, $\eta^2 = .11$, respectively. Children overall gave positive ratings to truth-telling and negative ratings to lie-telling. Also, children's ratings changed with age and setting. The mean effects were qualified by significant Age × Truthfulness and Truthfulness × Setting interactions, $F(2, 117) = 3.57$, $p < .05$, $\eta^2 = .06$, and $F(1, 117) = 106.95$, $p < .0001$, $\eta^2 = .49$, respectively. These effects were qualified by a significant Age × Truthfulness × Setting interaction, $F(2, 117) = 5.15$, $p < .001$, $\eta^2 = .08$. Simple effect tests with Sidak adjustments for multiple comparisons were performed to examine the significant three-way interaction. This effect was because of the fact that 7-year-olds were significantly different from 9- and 11-year-olds, whereas the two older groups' ratings of blunt truth-telling did not differ from each other. Seven-year-olds

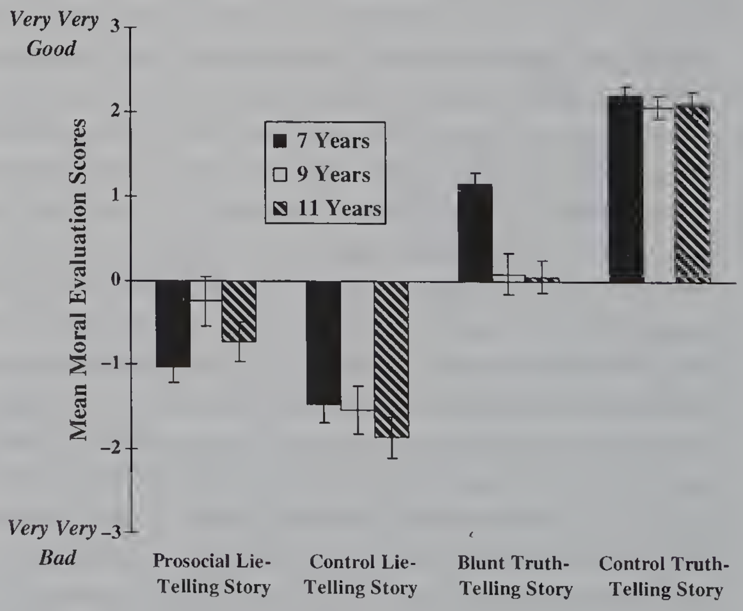


Figure 1. Means and standard errors of children's positive and negative evaluations of the story characters' statements by age.

gave significantly more positive ratings to blunt truth-telling than older children. In contrast, no age differences were found for control truth- and lie-telling and prosocial lie-telling. Overall, they gave very positive ratings to control truth-telling and very negative ratings to control lie-telling. They gave significantly less negative ratings to prosocial lie-telling than control lie-telling.

With regard to children's justifications of their evaluations, overall, participants did not provide a diverse array of justifications. For the control lie- and truth-telling stories in which the protagonist receives a desirable gift, all children justified their ratings to the former by referring to the negativity of dishonesty and the latter by referring to the positivity of honesty (the interrater reliability was 100%). Owing to the ceiling effect, the explanation data were not analyzed further. For the prosocial lie-telling and blunt truth-telling stories, the interrater reliability was also high, $\kappa s = .93$ and .85, respectively. The children's justifications fell into three types of responses: politeness justification, honesty justification, and both justification (see Table 1). No children gave ''I don't know'' responses.

A multinomial logistic regression analysis was performed with the age as the predictor and the type of justifications (politeness, honest, and both) for the prosocial lie-telling story as the predicted. The model (i.e., the age effect) was significant, $\chi^2(4, N = 120) = 19.41, p < .01$, Nagelkerk $R^2 = .17$. A priori comparisons with the honesty justification type as the reference for the predicted variable category and 7-year-olds as the reference for the predictor variable category were performed. As shown in Table 1, 9- and 11-year-olds used significantly more the ''both'' justification for their moral evaluations than did the 7-year-olds, $\beta = 1.82$, Wald = 6.18, $p < .01$, odds ratio = 9.90, and $\beta = 2.28$, Wald = 13.02, $p < .001$, odds ratio = 9.80, respectively. Thus, 9-year-olds were 6.18 times, and 11-year-olds were 9.8 times, more likely to use the both justification than 7-year-olds. No significant age differences were found for the politeness justification.

Table 1

Percent (Frequency) of Types of Justifications Made by Children at Each Age Group About Their Moral Judgments

	Politeness	Honesty	Both
Prosocial lie-telling			
7 years	20 (8)	70 (28)	10 (4)
9 years	20 (8)	43 (17)	38 (15)
11 years	10 (4)	38 (15)	53 (21)
Blunt truth-telling			
7 years	30 (12)	60 (24)	10 (4)
9 years	33 (23)	48 (19)	20 (8)
11 years	15 (6)	28 (11)	58 (23)

Another multinomial logistic regression analysis was performed with the age as the predictor and the type of justifications (politeness, honesty, and both) for the blunt truth-telling story as the predicted. The model (i.e., the age effect) was significant, $\chi^2(4, N = 120) = 24.68, p < .001$, Nagelkerk $R^2 = .21$. A priori comparisons with the honesty response type as the reference for the predicted variable category and 7-year-olds as the reference for the predictor variable category were performed. As shown in Table 1, 11-year-olds, but not 9-year-olds, used significantly more of both justifications for their moral evaluations than did the 7-year-olds, $\beta = 2.53$, Wald = 15.02, $p < .001$, odds ratio = 12.55. In contrast, no significant age differences were found for the politeness justification.

Two hierarchical linear regression analyses were performed with children's moral evaluation scores of prosocial lies and blunt truths as the predicted variable and age entered first as a predictor followed by the type of justifications, which was dummy coded as Predictor 1 (both vs. the other two responses) and Predictor 2 (honesty vs. the other two responses). After partialing out the effect of age, the type of justifications was significantly related to children's moral evaluation scores for prosocial lies and blunt truths, $F_{change}(2, 116) = 30.67$, $R^2_{change} = .34$, $p < .001$, and $F_{change}(2, 116) = 15.69$, $R^2_{change} = .20$, $p < .001$, respectively. Regardless of age, those children who used the both response to justify their ratings tended to give positive ratings to prosocial lies ($M = 0.45$, $SD = 1.54$) and negative ratings to blunt truths ($M = -0.49$, $SD = 1.27$). Those children who used the politeness response to justify their ratings gave moderately negative ratings to both prosocial lies ($M = -0.30$, $SD = 1.81$) and blunt truths ($M = -0.10$, $SD = 1.38$). In contrast, those children who used the honesty response to justify their ratings gave highly negative ratings to prosocial lies ($M = -1.53$, $SD = 0.81$) but highly positive ratings to blunt truths ($M = 1.31$, $SD = 1.50$).

Children's Truth- or Lie-Telling Behaviors in the Undesirable Gift Procedure

Children were coded as lie-tellers if they responded to their own teacher that they liked the gift she gave them (i.e., the Liking Question 1) but told the experimenter that they really did not like the gift (i.e., the Liking Question 2). If they responded that they did not like the gift to both questions, they were coded as blunt truth-tellers. Forty percent of 7-year-olds, 50% of 9-year-olds, and 60% of 11-year-olds were categorized as lie-tellers. A binary logistic regression analysis was performed with the age as the predictor and the child behavior (lie-telling vs. blunt truth-telling) as the predicted. The model (i.e., the age effect) was significant, $\chi^2(2, N = 120) = 10.72$, $p < .001$, Nagelkerk $R^2 = .12$. A priori comparisons with the 7-year-olds as the reference showed that 11-year-olds, but not 9-year-olds, differed significantly from 7-year-olds in terms of their tendency to tell a lie in the politeness situation, $\beta = 1.62$, Wald = 9.02, $p < .05$, odds ratio = 5.17 (11-year-olds were statistically 5.17 times more likely to lie than 7-year-olds).

With regard to children's responses to the question as to why they told the truth or lied, among the truth-tellers, their responses were highly similar. All of them either cited that they must tell the truth or simply said that they just did not like the gift. They were all coded as blunt truth-tellers by the two independent coders. Among the lie-tellers, one group of children were coded as prosocial lie-tellers because they justified their lying by referring to the need to be polite or the potential negative impact of their statement on the lie recipient if truth was told. Another group of lie-tellers (12 children in total) justified their lies by referring to the need to avoid negative consequences to themselves (e.g., the teacher being angry at them if truth be told). Yet another group of lie-tellers stated that they deserved the gift even though they did not like it (5 children in total). Another group (5 children in total) simply said: "I don't know." The rest gave miscellaneous responses unrelated to politeness (e.g., "I can find another use for it"). These children who gave these nonprosocial justifications were coded as nonprosocial lie-tellers. The intercoder reliability was high, $\kappa = .88$. Figure 2 shows the percentages of the three groups of children per age group.

A multinomial logistic regression analysis was performed with age as the predictor and the child behavior (blunt truth-telling, prosocial lie-telling, and nonprosocial lie-telling) as the predicted. The model (i.e., the age effect) was significant, $\chi^2(4, N = 120) = 11.27$, $p < .05$, Nagelkerk $R^2 = .10$. A priori comparisons were performed with 7-year-olds as the reference for the predictor variable and the blunt truth-tellers as the reference for the predicted variable. As shown in Figure 2, the proportion of the children who told a nonprosocial lie did not change with age. However, as age increased, more children told prosocial lies. Among the prosocial lie-tellers, 11- but not 9-year-olds differed significantly from 7-year-olds in terms of their tendency to tell a lie for prosocial purposes, $\beta = 1.67$, Wald = 4.13, $p < .05$, odds ratio = 4.75. That is, 11-year-olds were statistically 4.75 times more likely than 7-year-olds to tell a prosocial lie.

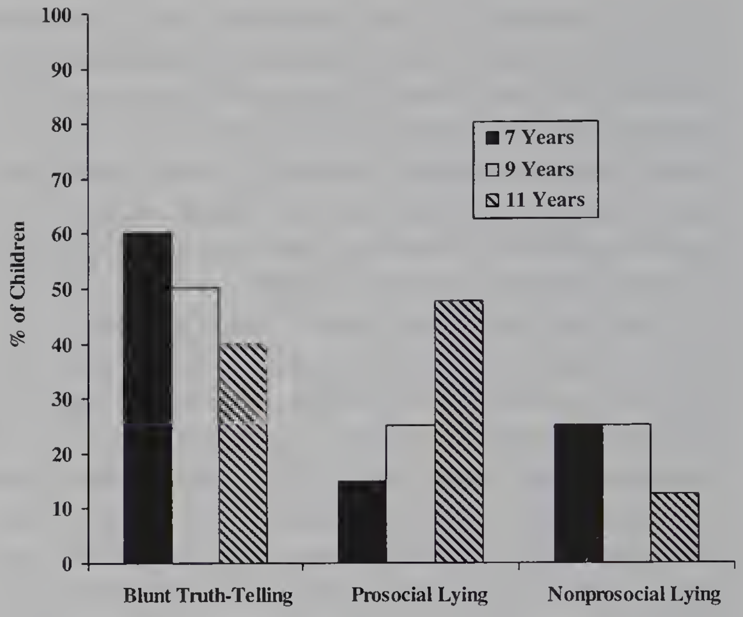


Figure 2. Percentage of prosocial lie-tellers, nonprosocial lie-tellers, and blunt truth-tellers by age.

Relations Between Children's Behavior and Moral Judgments

A 3 (age: 7, 9, and 11 years) × 3 (child behavior: prosocial lie-telling, nonprosocial lie-telling, and blunt truth-telling) ANOVA was performed on children's moral evaluation scores of the prosocial lie-telling story (see Table 2). Only the child behavior effect was significant, $F(2, 111) = 3.76$, $p < .05$, $\eta^2 = .06$. A priori contrast with prosocial lie-telling as the reference showed that both blunt truth-telling and nonprosocial lie-telling children's moral evaluation scores were significantly more negative than those of the prosocial lie-telling children ($ps = .025$ and $.013$, respectively).

Another 3 (age: 7, 9, and 11 years) × 3 (child behavior: prosocial lie-telling, nonprosocial lie-telling, and blunt truth-telling) ANOVA was performed on children's moral evaluation scores of the blunt truth-telling story. The effects of both age and child behavior were significant, $F(2, 111) = 4.78$, $p < .05$, $\eta^2 = .08$, and $F(2, 111) = 8.06$, $p < .01$, $\eta^2 = .13$, respectively. For the significant age effect, a priori contrast with 7-year-olds as the reference

Table 2
Means (Standard Deviations) of Children's Positive or Negative Evaluations of the Story Characters' Statements by Age and Behavioral Type (Minimum Score = −3, Maximum Score = 3)

Behavior type	Prosocial lie-telling story	Control lie-telling story	Blunt truth-telling story	Control truth-telling story
Blunt truth-tellers				
7 years	−1.29 (1.12)	−1.58 (0.72)	1.25 (1.29)	2.13 (0.80)
9 years	−0.25 (1.77)	−1.45 (1.50)	0.40 (1.50)	2.20 (0.77)
11 years	−0.87 (1.09)	−1.75 (1.07)	0.19 (1.56)	2.37 (0.96)
Prosocial lie-tellers				
7 years	0.33 (1.51)	−1.00 (0.63)	0.83 (1.47)	2.17 (0.75)
9 years	0.00 (2.16)	−1.00 (2.00)	−1.10 (1.29)	1.70 (1.16)
11 years	−0.37 (1.50)	−1.68 (1.29)	−0.63 (1.01)	1.68 (1.00)
Nonprosocial lie-tellers				
7 years	−1.20 (0.63)	−1.50 (0.97)	1.10 (1.29)	2.30 (0.68)
9 years	−0.50 (1.90)	−2.20 (0.79)	0.60 (2.22)	2.10 (0.74)
11 years	−1.60 (2.19)	−2.80 (0.45)	2.20 (1.30)	2.60 (0.55)

showed that 11-year-olds', but not 9-year-olds', moral evaluation scores of the blunt truth-telling story were significantly less positive than those of the 7-year-olds ($p < .05$). For the significant child behavior effect, a priori contrast with prosocial lie-telling as the reference showed that both blunt truth-telling and nonprosocial lie-telling children's moral evaluation scores were significantly more positive than those of the prosocial lie-telling children ($ps < .001$ and .01, respectively). In other words, those children who told a prosocial lie also tended to view blunt truth-telling in less a positive light than the other two groups of children.

To examine the relation between children's moral evaluation justifications (both, honesty, and politeness) for the prosocial lie-telling story and the child behavior (blunt truth-telling, prosocial lie-telling, and nonprosocial lie-telling), a multinomial logistic regression analysis was performed with the child behavior as the predictor and the type of justifications as the predicted. The model (i.e., the child behavior effect) was significant, $\chi^2(4, N = 120) = 15.36$, $p < .01$, Nagelkerk $R^2 = .14$. A priori comparisons with prosocial lie-telling as the reference for the predictor and the honesty justifications as the reference for the predicted were performed. Those children who gave the both justification were significantly (8.84 times) more likely to tell prosocial lies than nonprosocial lies, $\beta = 2.18$, Wald = 9.02, $p < .01$, odds ratio = 8.84, and (3.13 times) more likely to tell prosocial lies than the blunt truth, $\beta = 1.14$, Wald = 5.85, $p < .05$, odds ratio = 3.13. In other words, children who considered both the honesty and politeness issues when judging the prosocial lie story were more inclined to tell a prosocial lie than those who focused solely on either honesty or politeness. Consistent with this result, those children who gave the politeness justifications for their moral judgment scores of the prosocial lie-telling story were equally likely to tell prosocial lies, nonprosocial lies, or the blunt truth.

Another multinomial logistic regression analysis was performed with the child behavior as the predictor and the type of justifications for the blunt truth-telling story as the predicted. The model was also significant, $\chi^2(4, N = 120) = 31.72$, $p < .001$, Nagelkerk $R^2 = .26$. A priori comparisons with prosocial lie-telling as the reference for the predictor and the politeness justifications as the reference for the predicted were performed. Those children who gave the both justification to their moral judgment scores of the blunt truth-telling story were significantly (11.07 times) more likely to tell prosocial lies than nonprosocial lies, $\beta = 2.41$, Wald = 10.24, $p < .01$, odds ratio = 11.07, and (9.09 times) more likely to tell prosocial lies than the blunt truth, $\beta = 2.21$, Wald = 16.40, $p < .001$, odds ratio = 9.09. In other words, children who considered both the honesty and politeness issues when judging the blunt truth-telling story were more inclined to tell a prosocial lie than those who only focus on either honesty or politeness. Consistent with this result, those children who gave the politeness justifications to their moral judgment scores of the blunt truth-telling story were equally likely to tell prosocial lies or nonprosocial lies or the blunt truth.

Discussion

This study examined children's prosocial lie-telling behavior in a politeness situation and its relation to their moral understanding of lie- and truth-telling. Overall, almost all children were able to identify lies and the truth correctly in the story procedure. Contrary to our hypothesis, no age differences were found. Unlike Lee and Ross (1997) who tested participants 12 years and above, but similar to Bussey (1999) who tested younger children, all the children in our study identified untruthful statements told with the intent to help another individual as lies. Our findings taken together with those of Bussey (1999) suggest that young children do not adhere to the suggestion that prosocial untruthful statements may not be lies at all (Sweetser, 1987). However, it may be that as children develop into adolescence and adulthood their conceptions of prosocial lies change (Lee & Ross, 1997). Future studies are needed to examine this possible developmental shift in the conception of prosocial lies.

The moral evaluation results are consistent with and complement previous findings (Bussey, 1999; Walper & Valtin, 1992) that children's evaluations of prosocial lies become less negative between preschool and elementary school age. Children in this study, who were all elementary school age, gave less negative moral evaluations for prosocial lie-telling than for control lie-telling. They also gave less positive ratings to blunt truth-telling than control truth-telling. In addition, whereas children between 7 and 11 years of age similarly gave negative yet near-neutral ratings to prosocial lies, 7-year-olds were more likely to give positive ratings to blunt truth-telling than 9- and 11-year-olds whose ratings were at the neutral level.

This tendency to rate prosocial lie- and blunt truth-telling around the zero point (neither good nor bad) may reflect children's increasing awareness of the fact that politeness situations involve two contradictory sets of rules that must be weighed against each other. This is demonstrated by children's justifications of their ratings. Older children were significantly more inclined than younger children to raise the issues of honesty and politeness simultaneously when justifying their ratings of prosocial lies and blunt truths. Furthermore, children who recognized both issues to be applicable to the politeness situation were more inclined to let the concern for politeness override the need to be truthful. As shown by the results of the multiple regression analyses, regardless of age, those children who used the both responses to justify their ratings tended to rate prosocial lies positively and blunt truth negatively. In contrast, those who used the politeness justifications rated both prosocial lies and blunt truths moderately negatively, suggesting that they might be ambivalent about whether the concern for politeness should determine the appropriate course of action for the situation. Not surprisingly, children who used the honesty justifications gave prosocial lies highly negative ratings whereas blunt truths were given highly positive ratings.

In terms of children's actual behavior in the politeness situation, 40% of 7-year-olds, 50% of 9-year-olds, and 60% of 11-year-olds told a lie. They told the gift-giver that they liked the undesirable gift but later told the experimenter that they did not really like the gift. The age pattern is generally in line with the findings of Talwar, Murphy, et al. (2007), who also found a significant increase in lie-telling between 3 and 11 years of age in a similar undesirable gift situation. However, because Talwar, Murphy, et al. did not probe children's motives for lying, the extent to which their children truly told prosocial lies for prosocial purposes was not clear.

Unlike Talwar, Murphy, et al. (2007), this study questioned children's motives for lying. When children were asked about why they had lied to the teacher, a developmental difference emerged between younger and older children. Although most younger lie-tellers were motivated to lie for nonprosocial reasons such as to avoid negative consequences for themselves (e.g., the gift-giver being mad at them), the majority of older lie-tellers were motivated to lie for prosocial reasons such as avoiding hurting the feelings of the gift-giver. These findings suggest that although the tendency for children to tell a lie increases with age in the politeness situation, their motives for telling such lies also change with age, from more self-serving at the younger age to more prosocial at the older age. Thus, the significant age effect seen in this study and perhaps that in Talwar, Murphy, et al. reflects a developmental change in children's tendency to tell a truly prosocially motivated lie.

It should be noted that the rate of lying in this study (50%) was substantially lower than the rate of lying (about 80%) in situations where children tell lies to conceal their own transgression (e.g., Talwar & Lee, 2002a). The exact nature of this discrepancy is unclear. One possibility is that children's tendency to lie may be more significantly influenced by their self-interests than by the need to help others as suggested by Talwar et al. (2004). Consistent with this suggestion, they found that although most children aged between 3 and 11 years refused to lie to conceal their parents' transgression, when their own responsibility for the transgression was absolved, approximately half of the children would lie for their parents. This possibility is also in line with the present finding that more than half of the 11-year-olds and the majority of the younger children told the truth so that they would not miss the opportunity to obtain a desirable gift. Furthermore, even among the children who lied, more than a half of the 7- and 9-year-olds and 20% of the 11-year-olds lied for nonprosocial reasons.

Another major finding of our study is that children's actual lying or truth-telling behaviors in the politeness situation were consistently related to their moral evaluations and justifications of the prosocial lie-telling or blunt truth-telling stories. Prosocial lie-tellers were more likely to rate others' prosocial lie-telling more positively and blunt truth-telling more negatively than did nonprosocial lie-tellers and blunt truth-tellers. The same children also tended to mention both the politeness and honesty rules when justifying their evaluations of a story character's prosocial lie-telling or blunt truth-telling in a politeness situation. These results suggest that the prosocial lie-tellers were aware of both the need to be polite and to avoid hurting others as well as the need to be truthful. Nevertheless, they decided to be polite and therefore tell a prosocial lie themselves in the situation. These findings are inconsistent with the existing studies with children that involved the Kolhbergian moral dilemmas or antisocial lying. These studies tended to show either weak or no linkages between individuals' moral knowledge and moral action (Arnold, 1989; Blasi, 1980; Talwar et al., 2002, 2004; Thoma & Rest, 1986).

As mentioned earlier, the disconnection was likely because of the mismatch between children's interpretations of the two situations. One places them in an everyday situation in which they must choose an action. The other uses stories to depict a situation about which they must make a moral evaluation. Thus, to avoid this kind of mismatch and to reveal a relation between children's moral knowledge and action, we must directly assess children's interpretations of both the situations. By implementing such a strategy in this study, we were able to identify two different types of lie-tellers: prosocial and nonprosocial lie-tellers. Both types of children lied in the undesirable gift situation but for different reasons. Nonprosocial lie-tellers told a lie either out of fear or could not justify their lies and showed the classic dissociation between moral knowledge and action. They evaluated prosocial lies negatively and blunt truths positively but failed to act according to their moral knowledge. In contrast, prosocial lie-tellers told a lie for politeness purposes and were consistent in their knowledge and action because they not only told prosocial lies in the undesirable gift situation themselves but also evaluated such lies positively and blunt truths negatively.

It should be noted that these children were apparently aware that the concern for both politeness and honesty were applicable to the situation. However, the concern for politeness seemed to play a more significant role in prosocial lie-tellers' decisions to lie than the concern for honesty. Also, the former seemed to override the latter when prosocial lie-tellers made moral evaluations of prosocial lies and blunt truths. Blunt truth-tellers were also consistent between their moral knowledge and behavior as they evaluated prosocial lie-tellers highly negatively and blunt truth-tellers highly positively, and acted accordingly themselves. Blunt truth-tellers appeared to identify both the undesirable gift situation they themselves encountered and the same situation depicted in the stories as a situation calling for honesty and thus displayed consistency between moral knowledge and action.

The present assessment of children's justifications to their moral judgments and the reports of their motives for their own actions not only highlight the importance of seeking out such information for methodological purposes but also raise a theoretical question about children's interpretative theories of mind and action. Chandler and his colleagues (Chandler, Sokol, & Hallett, 2001; Chandler, Sokol, & Wainryb, 2000) have argued that children's beliefs about rightness (i.e., morality) should be studied together with their beliefs about mental life; the separation of the existing developmental research on the two issues is both incomplete and arbitrary because beliefs about rightness are intimately related to beliefs about mental states such as beliefs and intentions. This approach has received some empirical support. For example, Wainryb, Brehl, and Matwin (2005) examined children's narrative accounts and moral evaluations of their own interpersonal conflicts with others. They found that children gave qualitatively different interpretations when events were viewed from the victim's or perpetrator's perspectives, and hence different moral judgments. Although Wainryb et al. did not directly study children's actual moral behavior, they suggested that "by integrating, within the study of moral development, children's interpretations of the social interactions that are at the basis of moral thinking, this approach brings us a step closer to conceptualizing the study of children's moral behavior" (p. 80).

The findings of this study suggest that this interpretative approach may indeed be fruitful for research on moral development in general and the relation between children's moral judgment and actual action specifically (see also Arsenio & Lemerise, 2004). Although this study revealed a significant relation between children's evaluations of prosocial lies and blunt truths and their actual behavior, the relation was strengthened and clarified when both children's justifications about their evaluations and their own behavior were taken into consideration. In particular, the relation was strengthened when children showed an appreciation of two different perspectives based on the concerns for honesty and politeness. It should be noted, however, that this study did not explicitly ask about children's interpretations of the story characters' beliefs and intentions in the moral judgment task nor their interpretations of the politeness situation they themselves encountered (e.g., their own beliefs as a lie- or truth-teller and the experimenter as a lie- or truth-recipient). Future studies need to take a genuine interpretative approach to ask these specific questions, which may reveal an even closer relation between children's moral knowledge and behavior than what has been found in this study.

In summary, this study examined school-aged Chinese children's moral understanding of lie- and truth-telling and their actual lying behaviors in a politeness situation. It revealed that children at all ages categorized untruthful statements as lies and truthful statements as the truth regardless of whether they were told for prosocial or nonprosocial purposes. However, as age increased, children became increasingly less negative and more positive about others' lying in politeness situations, and they were more inclined to tell lies in such situations themselves. Contrary to previous findings (e.g., Talwar et al., 2002), children's moral knowledge about lying was significantly related to their actual behaviors.

References

Arnold, M. (1989, April). *Moral cognition and conduct: A quantitative review of the literature*. Paper presented at the Society for Research in Child Development, Kansas City, MO.

Arsenio, W. F., & Lemerise, E. A. (2004). Aggression and moral development: Integrating social information processing and moral domain models. *Child Development*, *75*, 987–1002.

Binet, A. (1896). Psychology of prestidigitation. *Annual Report of the Board of Regents of Smithsonian Institution, 1894*, 555–571.

Blasi, A. (1980). Bridging moral cognition and moral action: A critical review of the literature. *Psychological Bulletin*, *88*, 1–45.

Bok, S. (1978). *Lying: Moral choice in public and private life*. New York: Random House.

Broomfield, K. A., Robinson, E. J., & Robinson, W. P. (2002). Children's understanding about white lies. *British Journal of Developmental Psychology*, *20*, 47–65.

Bussey, K. (1992). Lying and truthfulness: Children's definitions, standards, and evaluative reactions. *Child Development*, *63*, 129–137.

Bussey, K. (1999). Children's categorization and evaluation of different types of lies and truths. *Child Development*, *70*, 1338–1347.

Cameron, C. A., Chen, S., & Lee, K. (2001, October). *Actions speak louder than words: Chinese adults' socialization of truth telling*. Poster presented at Moving Toward Moral Excellence Conference of the Association of Moral Education, Vancouver, Canada.

Chandler, M., Fritz, A. S., & Hala, S. (1989). Small-scale deceit: Deception as a marker of two-, three-, and four-year-olds' early theories of mind. *Child Development*, *60*, 1263–1277.

Chandler, M. J., Sokol, B. W., & Hallett, D. (2001). Moral responsibility and the interpretive turn: Children's changing conceptions of truth and rightness. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 345–365). Cambridge, MA: MIT Press.

Chandler, M. J., Sokol, B. W., & Wainryb, C. (2000). Beliefs about truth and beliefs about rightness. *Child Development*, *71*, 91–97.

Cole, P. (1986). Children's spontaneous control of facial expression. *Child Development*, *57*, 1309–1321.

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin*, *115*, 74–101.

Darwin, C. (1877). A biographical sketch of an infant. *Mind*, *2*, 285–294.

DePaulo, B. M., & Bell, K. L. (1996). Truth and investment: Lies are told to those who care. *Journal of Personality & Social Psychology*, *71*, 703–716.

DePaulo, B. M., & Kashy, D. A. (1998). Everyday lies in close and casual relationships. *Journal of Personality & Social Psychology*, *74*, 63–79.

Fu, G., & Lee, K. (2007). Social grooming in the kindergarten: The emergence of flattery behavior. *Developmental Science*, *10*, 255–265.

Goodman, G. S., Myers, J. E. B., Qin, J., Quas, J. A., Castelli, P., Redlich, A. D., et al. (2006). Hearsay versus children's testimony: Effects of truthful and deceptive statements on jurors' decisions. *Law and Human Behavior*, *30*, 363–401.

Grice, H. P. (1980). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Hartshorne, H., & May, M. S. (1928). *Studies in the nature of character: Vol. 1. Studies in deceit*. New York: Macmillan.

Helwig, C. C., & Turiel, E. (2002). Children's social and moral reasoning. In P. K. Smith & C. H. Hart (Eds.), *Blackwell handbook of childhood social development* (pp. 476–490). Malden, MA: Blackwell.

Huesmann, L. R. (1998). The role of social information processing and cognitive schema in the acquisition and maintenance of habitual aggressive behavior. In R. G. Geen & E. Donnerstein (Eds.), *Human aggression: Theories, research, and implications for social policy* (pp. 73–109). San Diego, CA: Academic Press.

Kant, I. (1949). On a supposed right to lie from altruistic motives. In L. W. Beck (Ed.), *Critical of practical reason and other writings* (pp. 346–350). Chicago: University of Chicago Press.

Krupfer, J. (1982). The moral presumption against lying. *Review of Metaphysics, 36,* 103–126.

Lakoff, R. (1973). The logic of politeness, or minding your P's and Q's. In *Papers presented at the Ninth Regional Meeting of the Chicago Linguistic Society* (pp. 292–305). Chicago: Chicago Linguistics Society.

Lee, K. (2000). The development of lying: How children do deceptive things with words. In J. W. Astington (Ed.), *Minds in the making* (pp. 177–196). Oxford, UK: Blackwell.

Lee, K., & Ross, H. (1997). The concept of lying in adolescents and young adults: Testing Sweetser's folkloristic model. *Merrill-Palmer Quarterly, 43,* 255–270.

Leekam, S. (1993). Children's understanding of mind. In M. Bennett (Ed.), *The development of social cognition: The child as psychologist* (pp. 26–61). New York: Guilford.

Lewis, M. (1993). The development of deception. In M. Lewis & C. Saarni (Eds.), *Lying and deception in everyday life* (pp. 90–105). New York: Guilford.

Lewis, M., Stanger, C., & Sullivan, M. W. (1989). Deception in 3-year-olds. *Developmental Psychology, 25,* 439–443.

Lyon, T. D. (2000). Child witnesses and the oath: Empirical evidence. *Southern California Law Review, 73,* 1017–1074.

Nucci, L. P., & Turiel, E. (2000). The moral and the personal: Sources of social conflicts. In L. P. Nucci, G. G. Saxe, & E. Turiel (Eds.), *Culture, thought, and development* (pp. 115–137). Mahwah, NJ: Erlbaum.

Nyberg, D. (1993). *The varnishing truth: Truth-telling and deceiving in ordinary life.* Chicago: University of Chicago Press.

Peskin, J. (1992). Ruse and representations: On children's ability to conceal information. *Developmental Psychology, 28,* 84–89.

Peterson, C. C. (1995). The role of perceived intention to deceive in children's and adults' concepts of lying. *British Journal of Developmental Psychology, 13,* 237–260.

Peterson, C. A., Peterson, J. L., & Seeto, D. (1983). Developmental changes in ideas about lying. *Child Development, 54,* 1529–1535.

Piaget, J. (1932). *The moral judgment of the child* (M. Gabain, Trans.). New York: Free Press.

Polak, A., & Harris, P. L. (1999). Deception by young children following noncompliance. *Developmental Psychology, 35,* 561–568.

Rest, J. R. (1986). *Moral development: Advances in research and theory.* New York: Praeger.

Saarni, C. (1984). An observational study of children's attempts to monitor their expressive behavior. *Child Development, 55,* 1504–1513.

Siegal, M., & Peterson, C. C. (1998). Pre-schoolers' understanding of lies and innocent and negligent mistakes. *Developmental Psychology, 34,* 332–341.

Smetana, J. G., & Turiel, E. (2003). Moral development during adolescence. In G. R. Adams & M. D. Berzonsky (Eds.), *Blackwell handbook of adolescence* (pp. 247–268). Malden, MA: Blackwell.

Sodian, B. (1991). The development of deception in young children. *British Journal of Developmental Psychology, 9,* 173–188.

St. Augustine. (1952). *Treaties on various issues.* Washington, DC: Catholic University of America Press.

Strichartz, A. F., & Burton, R. V. (1990). Lies and truth: A study of the development of the concept. *Child Development, 61,* 211–220.

Sweetser, E. E. (1987). The definition of lie: An examination of the folk models underlying a semantic prototype. In D. Holland (Ed.), *Cultural models in language and thought* (pp. 43–66). New York: Cambridge University Press.

Talwar, V., Gordon, H., & Lee, K. (2007). Lying in the elementary school years: Verbal deception and its relation to second-order belief understanding. *Developmental Psychology, 43,* 804–810.

Talwar, V., & Lee, K. (2002a). The development of lying to conceal a transgression: Children's control of expressive behavior during verbal deception. *International Journal of Behavioral Development, 26,* 436–444.

Talwar, V., & Lee, K. (2002b). Emergence of white-lie telling in children between 3 and 7 years of age. *Merrill-Palmer Quarterly, 48,* 160–181.

Talwar, V., & Lee, K. (2008). Social and cognitive correlates of children's lying behavior. *Child Development, 79,* 866–881.

Talwar, V., Lee, K., Bala, N., & Lindsay, R. C. L. (2002). Children's conceptual knowledge of lie-telling and its relation to their actual behaviors: Implications for court competence examination. *Law and Human Behavior, 26,* 395–415.

Talwar, V., Lee, K., Bala, N., & Lindsay, R. C. L. (2004). Children's lie-telling to conceal a parent's transgression: Legal implications. *Law and Human Behavior, 28,* 411–435.

Talwar, V., Murphy, S., & Lee, K. (2007). White lie-telling in children for politeness purposes. *International Journal of Behavioral Development, 31,* 1–11.

Thoma, S. J., & Rest, J. R. (1986). Moral judgment, behavior, decision making, and attitudes. In J. R. Rest (Ed.), *Moral development: Advances in research and theory* (pp. 133–175). New York: Praeger.

Turiel, E. (2002). *The culture of morality: Social development, context, and conflict.* Cambridge, UK: Cambridge University Press.

Wainryb, C., Brehl, B. A., & Matwin, S. (2005). Being hurt and hurting others: Children's narrative accounts and moral judgments of their own interpersonal conflicts. *Monographs of the Society for Research in Child Development, 70* (Serial No. 281).

Walker, L. J. (2002). The model and the measure: An appraisal of the Minnesota approach to moral development. *Journal of Moral Education, 31,* 353–367.

Walper, S., & Valtin, R. (1992). Children's understanding of white lies. In W. Winter (Series Ed.), R. J. Watts, S. Ide, & K. Ehlich (Volume Eds.), *Politeness in language: Studies in history, theory and practice* (pp. 231–251). *Trends in Linguistics: Studies and Monographs, 59*. Berlin: Mouton de Gruyter.

Wilson, A. E., Smith, M. D., & Ross, H. S. (2003). The nature and effects of young children's lies. *Social Development, 12*, 21–45.

Zeng, H. Y. (2004). *Effects of story context on children's white lie understanding*. Master of Arts dissertation, South China Normal University, P.R. China.

Child Development, March/April 2010, Volume 81, Number 2, Pages 597–615

Preschoolers' Implicit and Explicit False-Belief Understanding: Relations With Complex Syntactical Mastery

Jason Low
Victoria University of Wellington

Three studies were carried out to investigate sentential complements being the critical device that allows for false-belief understanding in 3- and 4-year-olds ($N = 102$). Participants across studies accurately gazed in anticipation of a character's mistaken belief in a predictive looking task despite erring on verbal responses for direct false-belief questions. Gaze was independent of complement mastery. These patterns held when other low-verbal false-belief tasks were considered and the predictive looking task was presented as a time-controlled film. While implicit (gaze) knowledge predicted explicit (verbal) false-belief understanding, complement mastery and cognitive flexibility also supported explicit reasoning. Overall, explicit false-belief understanding is complexly underpinned by implicit knowledge and input from higher-order systems of language and executive control.

According to de Villiers (2000), children must understand and reproduce complement syntax before achieving an understanding of false beliefs. Complements are statements embedded under propositional attitude verbs within a sentence. In a recent meta-analysis of studies on false-belief understanding, Milligan, Astington, and Dack (2007) uncovered a stronger direction of effect from measures of individual language ability to false-belief comprehension than the reverse. The largest effect size was observed for children's mastery of embedded complement structures, but this finding remains unclear for two reasons. First, the effect size may be overstated as only four studies in the meta-analysis assessed the unique role of sentential complements. Second, the correlation between false-belief reasoning with mastery of sentential complements did not differ significantly from correlations with other measures of language ability (e.g., synonym production; Perner, Stummer, Sprung, & Doherty, 2002). The relation between standard false-belief tasks with a variety of individual language skills may partly reflect implicit–explicit changes in representational format underpinning children's understanding of mental states. As such, we need to clarify the link between complement mastery and false-belief reasoning to better identify the constraints on individual language ability relating to mental state understanding.

Complementation Enables Cognition

de Villiers singled out children's mastery of the grammatical argument structure of mental state and communication verbs taking realis complements (clauses with overt present or past tense) with their associated marking of points of view as prerequisite for representing false beliefs (e.g., de villiers 2004, 2005; de Villiers & de Villiers, 2000, 2003). Consider the sentence: "Homer said *the cat yodeled*." While the complement embedded under the verb is false (the cat yodeled), it is still true that Homer said something. The complement structure also licenses the protagonist's terms of reference even when they are not ours: "Homer said the cat yodeled, but in fact it meowed." de Villiers contends that overt evidence of truth or falsity from the verb *say* is easier to conceptualize and may be a

---

This research was funded internally at Victoria University of Wellington by a School of Psychology Research Grant (SoP22817). The author is grateful to Janet Astington, Helena Gao, Sophie Jacques, Josef Perner, Ted Ruffman, and Karen Salmon for advice, and three anonymous reviewers for insightful and constructive comments on the original article. The author also thanks the children who generously participated in the research, and Steve Cochran and the many research assistants for their valuable help in collecting the data.

Correspondence concerning this article should be addressed to Jason Low, School of Psychology, Victoria University of Wellington, P.O. Box 600, Wellington 6140, New Zealand. Electronic mail may be sent to Jason.Low@vuw.ac.nz.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8102-0015

precursor to understanding sentential complements that include analogic mental state verbs (e.g., *think*).

de Villiers and Pyers (2002) found that memory for the content of complement sentences embedded under communication and mental state verbs at an earlier test point predicted later false-belief comprehension and not the reverse. Training on communication verb sentential complements alone can also improve children's false-belief performance (and not vice versa; Hale & Tager-Flusberg, 2003). Schick, de Villiers, de Villiers, and Hoffmeister (2007) also found predictive correlations linking grasp of communication verb (*told*) complement structure in American Sign Language with false-belief understanding among deaf children. A criticism of the memory for complements task is that even though children are simply asked to restate the false utterance of the protagonist, those utterances may cry out for an explanation in terms of the deceptive motive of the actor. However, Lohmann and Tomasello (2003) found that even training children to talk about objects using complement frames that do not highlight deceptive experiences can facilitate false-belief understanding. These findings fit with the strong view that "sufficient reasoning ability (for false-belief computation, prediction and explanation) is defined by having language of the appropriate degree of complexity" (de Villiers & de Villiers, 2003, p. 338).

Implicit False-Belief Knowledge

One theoretic complication to de Villiers' account of false-belief development remains relatively unexplored: children's implicit knowledge. Clements and Perner (1994) ran a change-in-location false-belief task whereby they videoed children's eye gaze in anticipation of the protagonist reappearing to look for the target object. Children between 2 years 11 months and 3 years 7 months looked more at the correct location despite providing an incorrect verbal false-belief judgment. Older children (by 4 years) matched their verbal response to an overt false-belief question with their visual orienting gesture. It is possible that young children possess an implicit understanding of false beliefs, and implicit knowledge of belief might precede and contribute to explicit understanding. The possibility of a distinction between implicit and explicit false-belief knowledge fits with nondevelopmental research. For example, blindsight patients show no awareness of objects that fall into the affected portion of their visual field and yet are able to accurately intuit features of the objects (Perner & Clements, 2000). On a cautionary note, however, Clements and Perner's (1994) anticipatory eye gaze results might instead index a weaker form of *explicit* knowledge.

One way to disestablish the "implicit" nature of early false-belief knowledge is to argue that young children's visual orienting could reflect less confident (but explicit) expectations. However, Ruffman, Garnham, Import, and Connolly (2001) found that the young children who gave incorrect verbal false-belief predictions did not even want to bet a modest number of counters on the location as indicated by their correct eye gaze. These results suggest that children are not conscious of the knowledge conveyed in their eye gaze.

A second objection is that anticipatory gaze may reflect a seeing = knowing rule (e.g., O'Neill, 1996). Garnham and Ruffman (2001) redesigned the predictive looking false-belief task to comprise three hiding locations wherein the protagonist first explored the middle hiding location before placing the object in the left-hand location. A second character then moved the object to the right-hand location. If predictive gaze is based on a sensitivity to the protagonist not knowing, children should look in anticipation equally often to the left and irrelevant middle locations. However, children looked in anticipation clearly to the left-hand location; anticipatory gaze appears to be based on implicit false-belief understanding.

A third objection stems from studies, involving deaf children, uncovering how performances on low-verbal and traditional verbal measures of false belief are correlated with complement mastery (e.g.,de Villiers & de Villiers, 2003; Schick et al., 2007). However, the critical issue may not be whether tasks are verbal or low verbal per se, but whether a task taps into implicit versus explicit knowledge. In the low-verbal object-hiding theory-of-mind task used by Schick et al., (2007), participants were instructed to point to which of two identical boxes contained a target object upon determining a helper's visual access to the hiding event. When children point, they are explicitly aware of having gestured, and this may not be equivalent to eye glances that are factuality implicit (Goldin-Meadow & Alibali, 1999). The low-verbal and verbal false belief tasks used by Schick et al. may have correlated with complement mastery because both methods tapped explicit knowledge.

Finally, Onishi and Baillargeon (2005) found that 15-month-olds showed increased looking times to an event when the actor's search behavior violated a false belief compared with when it fitted a false belief. These findings have been replicated with 13-month-olds even when similarities to last person-object-location encoding are controlled for (e.g., Surian, Caldi, & Sperber, 2007). In Clements and Perner's (1994) study, however, very young 2-year-olds did not show implicit understanding via anticipatory gaze. In the predictive looking task, the target object is present; 2-year-olds' knowledge of the actual location of the object may interfere with their implicit knowledge of where a protagonist with a false belief might look. This interference could also partly reflect young children's fledgling ability to inhibit or select attention away from a true-belief suggestion (e.g., Leslie, German, & Pollizi, 2005). Of course, this does not mean that the predictive looking task is not appropriate for tapping into implicit false-belief knowledge per se; it may simply mean that the task involves other demands that may swamp the rudimentary nature of 2-year-olds' false-belief sensitivity. Using a simpler version of the predictive looking task where the target object was removed altogether, Southgate, Senju, and Csibra (2007) found that 2-year-olds looked first (and longer) to the location where an actor was expected to search given a false belief. The developmental timeline suggests continuity in implicit false-belief understanding from 2 to 4 years of age and, further, that early knowledge may be formatted "in a rudimentary and implicit form" (Onishi & Baillargeon, 2005, p. 257). A step forward for theory building is to understand how implicit knowledge underpins explicit understanding, and how factors such as language ability make explicit knowledge more accessible.

Rise of Explicit Understanding

One view is that through the endogenous cyclical process of redescription, implicit representations are recoded into explicit ones (with the highest format providing access to declarative verbal reporting; Karmiloff-Smith, 1992). By hypothesis, variation in implicit social-perceptual understanding may reasonably correlate with the extent of explicit verbal false-belief reasoning. Continuity (i.e., a correlation in performance) between implicit and explicit false-belief knowledge is bolstered by certain lines of evidence. First, implicit knowledge is a precondition for receptivity to explicit false-belief training (e.g., Clements, Rustin, & McCallum, 2000). Second, other incipient theory of mind insights such as the duration of infants' correct looking during referential acts are correlated with higher level cognitive outcomes (e.g., Wellman, Phillips, Dunphy-Lelii, & LaLonde, 2004).

To provide a concrete mechanism for how changes in explicit reflection might occur, Zelazo (1994) suggested that redescription could also take place over the course of problem solving when children demonstrate cognitive flexibility or the ability to construct a higher order rule to simultaneously integrate multiple conflicting representations of a single event. Performance on the Dimensional Change Card Sort (DCCS) task is closely correlated to explicit false-belief task performance, independent of age and general intelligence (e.g., Frye, Zelazo, & Palfai, 1995; Müller, Zelazo, & Imrisek, 2005). Zelazo, Gao, and Todd (2007) explain that the two kinds of tasks may be related because both require cognitive flexibility in relating two incompatible perspectives (e.g., in the DCCS—how it is possible to sort the same cards by shape and then by color; in the representational change false-belief task—how it is possible to directly respond with a past or present self-perspective). Preschoolers may become more proficient at reasoning through both types of problems when they flexibly integrate opposing knowledge into an embedded hierarchical "if-if-then" rule. The link between cognitive flexibility and explicit false-belief understanding must, however, be qualified in several ways. First, studies by Carlson and Moses testing other aspects of executive control reveal a strong association between conflict inhibition tasks with standard verbal false-belief performance (e.g., Carlson & Moses, 2001; Carlson, Moses, & Claxton, 2004). Second, Perner and Lang (2000) pointed out that explicit verbal false-belief tasks can be passed without necessarily reformulating two separate rules in a hierarchical three-term fashion (e.g., if I am looking for the marble, then here; if Sally is looking for the marble then there). Third, the evidence on whether explicit mental state understanding or cognitive flexibility over rule switching emerges first is equivocal (e.g., Kloo & Perner, 2003).

Nonetheless, even views concentrating on the cognitive flexibility facet of executive control in relation to advancing mental state awareness maps onto the broader point made here that the grammar of complement constructions cannot be the critical device that allows for false-belief representation. Jacques and Zelazo (2005) have suggested that the relation between individual language ability and explicit false-belief answering can instead be meaningfully recast in terms of language working to service cognitive flexibility in promoting explicit perspective taking. To the extent that cognitive flexibility in rule use refers to linguistically mediated reasoning for promoting explicit control of thought and action (Carlson, 2003), then such a view even fits with research indicating that verbally labeling cards before sorting on the DCCS task can improve inhibitory control to support explicit attentional focus on multiple and conflicting perspectives (e.g., Kirkham, Cruess, & Diamond, 2003). One integrated view may be that, alongside an implicit understanding of false belief, skill sets that include individual language ability (e.g., complement mastery) and cognitive flexibility can help make emergent explicit understandings of false belief.

Only Tantalizing?

Unfortunately, studies emphasizing early implicit sensitivity to false-belief understanding remain tantalizing because they do not include any measure of children's individual language abilities (e.g., Clements & Perner, 1994; Clements et al., 2000; Ruffman et al., 2001). If it turns out that visual orienting does not uniquely tap implicit knowledge but is correlated with complement mastery, it is possible to maintain that certain syntactic structures provide the breakthrough for any false-belief understanding (e.g., de Villiers, 2000, 2004; de Villiers & de Villiers, 2003). If eye gaze is not correlated with mastery of complement structure, it would mean that although certain grammatical platforms may play a role in harnessing explicit levels of false-belief understanding that are more complex and abstract, complementation is unlikely to provide the conceptual foundation for children's metarepresentational ability (Surian et al., 2007). Implicit knowledge is also likely to contribute to explicit false-belief understanding. For the present investigation, two distinct but connected premises were submitted for analysis: (a) explicit false-belief understanding has distinctive roots in implicit knowledge (via anticipatory eye gaze) and (b) cognitive flexibility and complex language ability uniquely contribute to underpinning children's explicit false-belief understanding.

Study 1

The first study examined whether complement mastery was correlated with anticipatory eye gaze and verbal judgments on standard false-belief tasks. Based on studies maintaining that visual orienting and verbal answers can potentially tap into implicit and explicit false-belief understanding, respectively (e.g., Clements & Perner, 1994; Ruffman et al., 2001), two main hypotheses were submitted. First, early false-belief sensitivity was expected to show up on indirect anticipatory eye gaze but not on direct verbal answering. Second, eye gaze was hypothesized to be language independent, but anticipatory gaze and complement mastery were each expected to uniquely contribute to predicting explicit false-belief understanding.

Method

Participants. Data came from 24 participants: twelve 3-year-olds (*M* = 3.8, *SD* = 2.09 months, range = 3.5–3.10; 4 girls and 8 boys) and twelve 4-year-olds (*M* = 4.7, *SD* = 2.62 months, range = 4.1–4.10; 6 girls and 6 boys). A total of 30 children were interviewed but 6 children (four 3-year-olds and two 4-year-olds) were not included in the final analysis as they failed to answer control questions in the various false-belief tasks correctly and did not sit still during the eye gaze task to allow examination of predictive looking. All participants in this study (and also in Studies 2 and 3) were Pakeha (New Zealanders of European descent), spoke English as their first language, and were from middle-class families.

Procedure. All children were individually interviewed in a quiet room at their kindergarten, and each of the 24 data participants took part in four 10- to 15-min fixed-order sessions with an interval of about 1 week between each. In Session 1, the eye-gaze false-belief task was presented. In Session 2, the Peabody Picture Vocabulary Test (PPVT)–III was administered. In Session 3, the Matrices and complement mastery tasks were presented. In Session 4, the unexpected contents, deceptive appearance, and change-in-location false-belief tasks were presented.

Verbal and nonverbal ability. Children's verbal ability was assessed using the PPVT–III (Dunn & Dunn, 1997). Nonverbal ability was measured using Raven's Colored Matrices (Raven, 1993). Raw scores of both measures were used.

Complementation task. The task measuring memory for complements embedded under a communication verb (*say*) was derived from Hale and Tager-Flusberg (2003). Each story involved different characters saying they did one activity when they were doing something else. Stories were worded to avoid any deceptive context or invitations to consider the motives of the false utterance. The order in which the information about what the central character said and what he or she did was counterbalanced across trials. Each trial was accompanied by two drawings, and at the end of each story participants were asked what the central character said. In one trial the experimenter presented a picture of a boy by himself baking cookies and said, "James was baking some cookies for his friend Lucy." The experimenter replaced the first picture with a second one showing the boy and his friend without any action and said, "Lucy came along and asked him what he was doing. James said, "I'm baking some bread." Participants were asked the test question, "What did James say he was baking?" The remaining three trials were about a girl saying she was cutting paper when she was really cutting her hair, a boy saying he kissed his mother when he really kissed Big Bird, and a girl saying she played with a cat when she really played with a dog. Each correct answer received 1 point for a total score of 4.

Eye gaze. Ruffman et al.'s (2001) predictive gaze task was used to measure implicit false-belief understanding. A 3D model (80 cm length × 35 cm width × 50 cm height) of a house was displayed on a table. At the front there was a ladder going to the top and a slide exit on either side (red at left, green at right). Slide exits at the front were separated by a distance of 70 cm and were adjacent to the slide entrances hidden at the back of the model. A red and green box (7 cm × 7 cm × 7 cm) was positioned 2 cm from either side of the respective slide exits. Toy objects were used to enact the event. The child sat on the floor in front of the model, with a camera above it recording eye gaze (see Figure 1). An assistant hidden behind the model house helped the experimenter with the staging of the event.

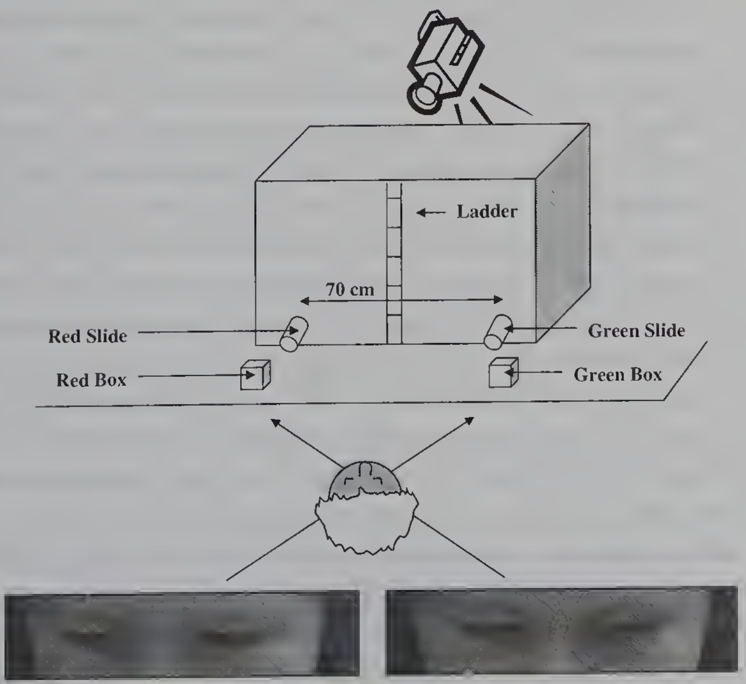


Figure 1. Eye gaze apparatus setup with illustrative frames of gaze direction.

The experimenter first demonstrated to participants that when the central protagonist wanted to look in the red box he would come down the red slide, and when he wanted to look in the green box he would come down the green slide. All children remembered where the protagonist would emerge from to look in the respective boxes. For the false-belief story, the experimenter enacted a narrative whereby the protagonist hid his toy car in the red box and then went to sleep behind the house (the assistant's hand appeared from behind to remove the protagonist upon climbing to the top of the house). Then the experimenter enacted the next episode where another character came along and moved the toy from the red to the green box and left the scene. The experimenter then sat behind the child. The assistant hidden behind the model house then announced that the central character was waking up and wanted to go play with his toy. The experimenter then wondered aloud to elicit anticipatory gaze, "I wonder which slide [protagonist] will come down to look for his car." Four seconds later, children were asked an explicit false-belief question: "Which slide will [protagonist] come down to look for his car?" A true-belief control story with the protagonist (a different character) witnessing the target (a different object) changing locations from the red to green container was also enacted. False- and true-belief trials were counterbalanced. Participants passed the true-belief explicit question and looked longer to the green slide location, providing confidence that gaze on the false-belief trial was not because of participants retracing story events and looking to where the protagonist first put the object. All participants passed control questions asked after the standard explicit false- and true-belief questions: reality (where is the [object] now?), memory (where did [protagonist] put his [object] in the beginning?), and perceptual access (did [protagonist] see [displacer] move the [object]?).

Videotapes of participants' eye gaze for the false-belief trial over a 4-s period following the "I wonder" prompt were analyzed to determine the number of frames during which children looked to a given location when anticipating the character's descent. Miyuki and Chisato (2006) pointed out that the whites of human eyes produce strong contrast with the pupils when looking left or right, and in that sense, detection of gaze direction can be relatively straightforward. Using VideoReDo Plus, one rater coded all tapes in slow motion and scored gaze direction on a frame-by-frame basis (looking times were accurate to 0.04 s; 25 frames/s). Another rater coded 20% of the videotapes, and raters agreed 100% of the time as to whether or not the child was looking more to the correct location. The gaze task captures rich variance in children's implicit false-belief knowledge via latency of visual orienting over 4 s; explicit understanding, however, was only tapped by verbal answers to a single direct false-belief question. To ensure sufficient variance in the measurement of explicit knowledge, three other direct false-belief tasks were included.

Unexpected contents and deceptive appearance tasks. Single trials of the unexpected contents and deceptive appearance tasks were given (Bialystok & Senman, 2004; Gopnik & Astington, 1988). For unexpected contents, participants were shown an M&Ms tube and asked what it contained (all said ''M&Ms''). They were then shown that it contained a spoon instead of M&Ms. Participants were asked about their own false belief (When you first saw this tube, what did you think was inside?) as well as to predict what an out of sight puppet (Ernie) would think was in the tube (When Ernie sees this tube, what will he think is inside?). Participants were then asked the reality control question (What is inside the tube right now?). For deceptive appearance (involving a sponge rock), participants were shown the object placed on a table and asked, ''What is this?'' All children answered correctly (rock or stone). Next the object's hidden property was revealed and two direct false-belief questions of the object's appearance were asked: ''What did you think this was when you first saw it?'' and ''Ernie did not hear or see what we were doing; what will Ernie think this is?'' Participants were then asked the reality control question (What is this really?).

Change-in-location task. The unexpected contents and deceptive appearance tasks contributed a maximum of 2 points each to explicit understanding (two direct false-belief test questions in each context). Verbal answers to the single direct question in Ruffman et al.'s (2001) gaze task contributed 1 point to explicit false-belief understanding as presented in a change-in-location context; thus, a single trial of Baron-Cohen, Leslie, and Frith's (1985) Sally-Ann change-in-location task was also administered. Toy characters were used and the hiding locations employed a basket and a jug with a small plastic fruit serving as the target object. The explicit test question involved children answering where the central protagonist (Sally) would first look for the (displaced) object. Then participants were asked two control questions: ''Where is the fruit now?'' and ''Where did Sally put the fruit?''

Overall, six explicit false-belief questions were asked and their contexts (unexpected contents, deceptive appearance, and change in location) contributed equal subscores to explicit false-belief answering. All participants answered the associated control questions correctly. A composite explicit false-belief score was calculated for each participant by adding the points across all direct test questions yielding a maximum score of 6 ($\alpha = .82$).

Results and Discussion

Preliminary analyses revealed that there were no gender effects in Study 1 (or in Studies 2 and 3), and including gender in the regression models did not affect the results. Consequently, gender was not included in all analyses reported here.

Eye gaze. The location of children's initial look following the anticipatory ''I wonder'' prompt was analyzed first. Significantly more children showed correct first looks ($N = 21/24$; 87%) than expected by chance based on the binomial distribution ($p < .001$). As a second approach to the data, eye gaze was assessed in terms of the total amount of time spent looking at the correct and incorrect locations. There was no significant age group difference for the amount of time spent looking at the correct location or the incorrect location (see Table 1 for mean scores). Nonetheless, collapsed across age groups, children spent a longer total amount of time looking at the correct location ($M = 1.45$ s, $SD = 0.78$) than the incorrect location ($M = 0.83$, $SD = 0.69$), $F(1, 23) = 8.02$, $p < .01$, $\eta_p^2 = .26$. Finally, an analysis of participants' visual orienting preferences relative to their ability to give a correct verbal answer on the single explicit false-belief question for the gaze task was carried out. A conservative metric was used to capture the *degree* of correct preferential looking whereby participants were classified as passing eye gaze only if the looking time *difference* for the false-belief trial ([looking at correct location] − [looking at incorrect location]) was at least +0.15 s. There was a significant link between anticipatory gaze and giving a correct verbal answer for the explicit question, $\chi^2(1, N = 24) = 4.51$, $p < .05$, Cramer's $\phi = .43$. Ten children (42%) passed eye gaze and correctly answered the single explicit false-belief question while 5 children failed both (21%). Critically, 37% ($n = 9$) of

Table 1

Mean Performance on Various Measures of Interest (Standard Deviations Shown in Parentheses)

Measure	3-year-olds	4-year-olds	Total	Range	Age difference
Study 1 ($N = 24$)					
Eye gaze (s)					
At correct location	1.22 (0.83)	1.67 (0.68)	1.45 (0.78)	0–2.52	*ns*
At incorrect location	0.77 (0.58)	0.88 (0.80)	0.83 (0.69)	0–2.68	*ns*
Matrices	10.42 (3.96)	9.42 (4.74)	9.92 (4.30)	2–18	*ns*
Composite explicit false-belief	2.25 (2.09)	3.92 (1.66)	3.08 (2.04)	0–6	$F(1, 22) = 4.63^{*}, \eta_p^2 = .17$
PPVT	57.75 (9.32)	66.17 (9.64)	61.96 (10.23)	50–85	$F(1, 22) = 4.72^{*}, \eta_p^2 = .18$
Complement mastery	1.83 (1.11)	2.83 (0.83)	2.33 (1.09)	0–4	$F(1, 22) = 6.19^{*}, \eta_p^2 = .22$
Study 2 ($N = 36$)					
Eye gaze (s)					
At correct location	1.31 (1.05)	1.61 (0.76)	1.46 (0.92)	0–3.68	*ns*
At incorrect location	0.75 (0.72)	0.53 (0.53)	0.64 (0.63)	0–2.88	*ns*
Matrices	10.00 (3.64)	11.78 (4.14)	10.89 (3.95)	3–20	*ns*
Composite explicit false-belief	0.61 (0.57)	1.43 (0.66)	1.02 (0.73)	0–2	$F(1, 34) = 15.95^{***}, \eta_p^2 = .32$
PPVT	55.39 (16.96)	68.94 (15.05)	62.17 (17.23)	18–93	$F(1, 34) = 6.43^{*}, \eta_p^2 = .16$
Complement mastery	1.89 (1.45)	2.83 (1.15)	2.36 (1.38)	0–4	$F(1, 34) = 4.68^{*}, \eta_p^2 = .12$
DCCS	1.11 (0.76)	1.94 (0.64)	1.53 (0.81)	0–3	$F(1, 34) = 12.71^{**}, \eta_p^2 = .27$
Study 3 ($N = 42$)					
Eye gaze (s)					
At correct location	1.77 (0.88)	1.98 (0.79)	1.87 (0.83)	.20–3.60	*ns*
At incorrect location	0.59 (0.50)	0.66 (0.68)	0.63 (0.59)	0–2.28	*ns*
Matrices	13.05 (2.51)	14.05 (2.94)	13.55 (2.75)	10–19	*ns*
Composite explicit false-belief	2.10 (1.67)	3.38 (1.96)	2.74 (1.91)	0–6	$F(1, 40) = 5.23^{*}, \eta_p^2 = .12$
PPVT	40.38 (13.69)	57.57 (17.23)	49.00 (17.90)	22–83	$F(1, 40) = 12.37^{**}, \eta_p^2 = .24$
Complement mastery	1.61 (1.20)	2.57 (1.29)	2.10 (1.32)	0–4	$F(1, 40) = 6.14^{*}, \eta_p^2 = .13$
DCCS	1.33 (0.86)	2.38 (0.60)	1.86 (0.90)	0–3	$F(1, 40) = 21.32^{***}, \eta_p^2 = .35$

Note. PPVT = Peabody Picture Vocabulary Test; DCCS = Dimensional Change Card Sort.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

children passed eye gaze while erring on their verbal answers to the explicit false-belief question but no participant showed the reverse relation.

Group differences on explicit false-belief and other measures. Mean scores on each of the other measures are also reported in Table 1. A one-way ANOVA indicated that 4-year-olds had a higher aggregate explicit false-belief score than 3-year-olds. Similar analyses also revealed significant age group differences in PPVT performance and complement mastery. There was no significant age group difference in general nonverbal ability on the Matrices task.

Correlations and hierarchical regressions. For the remaining analyses in Study 1, to examine empirical connections between implicit and explicit false-belief understanding, the difference in looking time variable was used as the candidate metric for reflecting implicit knowledge. The reasoning is as follows. First, a looking time difference metric converges with approaches used in other relevant developmental literature, for instance, gaze following and social attention, where correlations between visual orienting durations and higher cognitive task performance are computed and interpreted (e.g., Brooks & Meltzoff, 2005; Wellman et al., 2004). Other metrics such as the correct first look, while important for determining that children's predictive looking behaviors do not reflect chance responding, do not yield sufficient variation to avoid spuriously low correlations with explicit false-belief understanding (or complement mastery; cf. Bornstein & Sigman, 1986). Finally, looking-time difference provides a clearer index of a state of mind as reflected in the degree of preference in children's visual orienting responses (Ruffman, 2000) and, further, duration of difference in looking time by yielding variation fits with arguments that there is real, though subtle, growth in implicit knowledge (Munakata, McClelland, Johnson, & Siegler, 1997). Full and partial correlations are reported in Table 2 (top row). Full correlations indicated that explicit false-belief understanding (composite score), aside from being related to chronological age, was related to

Table 2

Bivariate and Full Partial Correlations (in Parentheses), Among Variables of Interest in Studies 1 (N = 24), *2* (N = 36), *and 3* (N = 42)

Variable	1	2	3	4	5	6
Study 1						
Age	—	−.10	.45*	.67***	.61**	.24
Matrices		—	−.22	.21	.12	−.06
PPVT			—	.38	.51*	.23
Complements				—	.71*** (.44*)	.13 (−.06)
Explicit false belief					—	.54** (.53**)
Eye gaze						—

Variable	1	2	3	4	5	6	7
Study 2 (above diagonal) and study 3 (below diagonal)							
Age	—	.26	.47**	.41*	.43**	.56***	.27
Matrices	.20	—	.30	.26	−.02	.12	.37*
PPVT	.58***	.20	—	.35*	.08	.41*	.23
Complements	.31*	−.04	.55***	—	.47** (.42*)	.58*** (.46**)	.31 (.19)
Cognitive flexibility	.50**	−.15	.34*	.48** (.38*)	—	.49** (.38*)	.04 (−.04)
Explicit false belief	.42**	−.07	.49**	.60*** (.40*)	.53*** (.37*)	—	.48** (.45**)
Eye gaze	.17	−.11	.21	.13 (−.01)	.23 (.12)	.55*** (.50**)	—

Note. Full partial correlations control for age alongside scores on the PPVT and Matrices. PPVT = Peabody Picture Vocabulary Test.
*$p < .05$. **$p < .01$. ***$p < .001$.

eye gaze and language ability (both general vocabulary and complementation). There was no significant bivariate correlation between eye gaze and complement mastery.

When the effects of chronological age and general developmental abilities were controlled, explicit false-belief reasoning was related to eye gaze and complement mastery. Partial correlation results supported the hypothesis that spontaneous anticipatory eye gaze marked implicit knowledge that was not significantly related to individual differences in complement mastery. There appear to be dissociations in preschoolers' false-belief knowledge between the behavioral and verbal levels. A hierarchical regression analysis was conducted to determine whether eye gaze and complement mastery each contributed unique variance in predicting explicit understanding (see Table 3).

Age, and PPVT and Raven's scores were entered as the first block of independent variables, and overall, they accounted for 49% of explicit false-belief variance. Analysis of regression coefficients revealed age, $\beta = .48$, $t(20) = 2.67$, $p = .015$, to be a focal predictor of explicit false-belief understanding. Eye gaze performance was entered in the second step, and variation in implicit understanding uniquely accounted for a further 14% of variance in explicit reasoning. Complement mastery was entered at the third step and accounted for further unique variance (11%) in explicit reasoning. Analysis of regression coefficients confirmed that, even at the third step, eye gaze performance continued to exert unique predictive value to explicit false-belief reasoning, $\beta = .41$, $t(18) = 3.34$, $p = .004$. Implicit knowledge and complement mastery appeared to uniquely scaffold preschoolers' explicit false-belief understanding.

Table 3

Results of Hierarchical Regressions for Variables Predicting Explicit False-Belief Performance in Study 1 (N = 24)

Variable	*B*	*SE B*	β	*T*	ΔR^2	Δ*F*	Sign. Δ*F*
Step 1							
Age	.16	.06	.48	2.67*	.49	6.51	.00
Matrices	.11	.08	.24	−0.45			
PPVT	.07	.04	.35	1.93			
Step 2							
Age	.14	.05	.41	2.59*	.14	7.43	.01
Matrices	.12	.07	.24	1.72			
PPVT	.06	.03	.29	1.82			
Eye gaze	.75	.27	.39	2.73*			
Step 3							
Age	.03	.06	.09	0.54	.11	7.89	.01
Matrices	.04	.06	.09	0.69			
PPVT	.04	.03	.20	1.45			
Eye gaze	.79	.24	.41	3.34**			
Complements	.93	.33	.50	2.81*			

Note. PPVT = Peabody Picture Vocabulary Test.
*$p < .05$. **$p < .01$.

Study 2

While the findings of Study 1 were consistent with theoretic predictions, three issues needed to be addressed further. First, the sample size was small, so it is important to replicate the findings. Second, de Villiers (e.g., de Villiers & de Villiers, 2003) argued that extant findings on anticipatory looking are difficult to align with evidence gained from deaf studies involving less verbal measures of false belief. If less or nonverbal tasks do not pattern after verbal ones, then deaf children should demonstrate increased skills on less verbal measures because such participants are more visually oriented, but this does not appear to be the case (e.g., Schick et al., 2007, using the object-hiding task; Woolfe, Want, & Siegal, 2002, using the thought-bubbles task). However, even with such less verbal false-belief tasks, children are still asked to make a direct prediction based on a protagonist's epistemic state, suggesting that those tasks essentially draw on explicit knowledge. In contrast, eye glances are not recorded in response to a direct question; rather, they are implicitly elicited in anticipation of a character's actions (Ruffman, 2000, 2004). To disentangle these ideas, Study 2 provided children a larger battery of false-belief measures that included the object-hiding and thought-bubbles tasks. Following appeals to implicit and explicit false-belief understanding, it was predicted that complement mastery would be correlated with explicit false-belief understanding (including performance on the object-hiding and thought-bubbles tasks) but not with implicit (anticipatory looking) knowledge. While the results of Study 1 supported the first theoretic submission that explicit false-belief decisions have roots in unique implicit knowledge and that grasp of complex language contributes to conscious understanding, I also submitted a second premise that explicit knowledge may be reconfigured at higher levels of disembodiment through cognitive flexibility and natural language. While implicit knowledge via eye gaze was predicted to correlate with explicit verbal false-belief reasoning (as in Study 1), DCCS performance and complement mastery were predicted to each correlate with explicit understanding.

Method

Participants. New data came from 36 participants: eighteen 3-year-olds (M = 3.6, SD = 2.38 months, range = 3.4–3.11; 5 girls and 13 boys) and eighteen 4-year-olds (M = 4.4, SD = 2.94 months, range = 4.1–4.10; 6 girls and 12 boys). A total of 43 children were tested but five 3-year-olds failed to answer control questions in the false-belief tasks correctly and two 3-year-olds did not attend the final testing session, and so those individuals were not included in the final analysis.

Procedure. Participants were individually tested in a room at their kindergarten across six 10- to 15-min sessions with an interval of 1 week between each. Dovetailing with Study 1, tasks were presented in the following order. In Session 1, the eye-gaze task was presented. In Session 2, the PPVT–III was administered. In Session 3, the Matrices and complement mastery tasks were given. In Session 4, the unexpected contents, deceptive appearance, and change-in-location false-belief tasks were presented. In Session 5, the DCCS and thought-bubbles tasks were given. In Session 6, the object-hiding task was presented.

DCCS (full version). Participants were shown two target cards (blue rabbit and red boat) attached to one of two target trays and told of two preswitch rules (e.g., by color) for sorting a series of bivalent cards (red rabbits and blue boats). After successfully completing a practice trial sorting two cards, participants started the preswitch phase (e.g., "If it's blue it goes here, and if it's red it goes there. This is a blue one, where does it go?"). No feedback was given. Participants needed to sort all six preswitch trials correctly in order to receive a score of 1 (Zelazo, 2006).

For the postswitch phase participants were told: "Now we are not going to play the color game anymore. We are going to play a shape game now. In the shape game, all rabbits go here and all boats go there. If it's a rabbit it goes here, and if it's a boat it goes there." The experimenter picked up a card and labeled it by its relevant dimension, and the child had to sort cards independently (e.g., "This is a rabbit, where does it go?"). Participants had to sort at least five of six trials to pass the postswitch phase (Zelazo, 2006).

The full DCCS task comes with a third phase (the border trials) that requires a higher level of cognitive flexibility. Some of the test cards were presented with black borders and the others were normal: "Now we are playing a different game. If there is a black line around the card like this one (red rabbit with a border), we play the color game. In the color game, all red ones go here and all blue ones go there. This card is red, so we put it over here. If there is no black line around the card, like this one (red rabbit without a border), we play the shape game. In the shape game, all rabbits go here and all boats go there. This one is a rabbit, so we put it here.'' This phase comprised 12 trials (three red rabbits and three blue boats without borders, three red rabbits and three blue boats with borders). On each trial, the rule was repeated. Children must sort at least 9 of 12 cards correctly to pass this phase (Zelazo, 2006).

The full DCCS was administered to capture continuous changes on cognitive flexibility over rule use (Hongwanishkul, Happaney, Lee, & Zelazo, 2005). Test cards were presented such that no more than two cards of the same type were presented consecutively. Each child's performance was scored as 0 (failed preswitch phase), 1 (passed pre- but failed postswitch phase), 2 (passed pre- and postswitch, but failed the border phase), or 3 (passed all phases; Zelazo, 2006).

False-belief tasks. With respect to Ruffman et al.'s (2001) eye-gaze task, all participants passed the true-belief control trial. One rater coded all tapes for the false-belief trial in slow motion and scored gaze direction on a frame-by-frame basis (looking times were accurate to 0.04 s). Another rater coded 20% of the videotapes, and raters agreed 100% of the time as to whether or not the child was looking more to the correct location. As in Study 1, single trials of the standard M&Ms, sponge-rock and Sally-Ann false-belief tasks were also administered. Overall, correct answers on the direct false-belief test questions for the task contexts of unexpected contents, deceptive appearance, and change in location each provided a subtotal score of 2, respectively. The new false-belief tasks were thought bubbles (Woolfe et al., 2002) and object hiding (Call & Tomasello, 1999).

False belief: Thought bubbles. A pretest was carried out to check that participants understood what thought bubbles meant. Two pictures were shown: one of boy taking a dog for a walk and the other of a boy thinking about a dog (with a thought bubble over his head containing a dog). The experimenter asked: ''Which boy is thinking about a dog?'' All children correctly pointed to the picture depicting a thought bubble. The task proceeded with two false-belief and two true-belief (control) trials (counterbalanced). An example false-belief trial showed a boy whose hand was reaching inside a partially shut kitchen cupboard. The experimenter said, ''Billy is reaching into the cupboard for a glass of water.'' Then the experimenter placed her palm over the protagonist's face and said, ''Let's cover Billy up so he does not see what is inside the cupboard.'' The cupboard door-flap was opened: ''Look, there is a mouse in the cupboard.'' (The picture shows Billy's hand reaching toward a mouse.) The flap was covered back, and the experimenter removed her palm covering the protagonist. She then presented another picture of Billy with an empty thought bubble over his head. A series of forced-choice picture items were shown underneath: a glass of water (false belief), a mouse (reality), an apple, and a bag (distracters). Children were instructed: ''Point to what Billy thinks is inside the cupboard.'' The other false-belief trial involved a boy who was fishing but whose line snagged a boot hidden behind reeds. True-belief trials involved a girl who sees a tall boy looking at her from behind a fence (the fence was removed to reveal a tall boy), and a boy walking along the beach who sees a fish tail from behind a bush (the bush was removed to reveal a fish). All participants passed the true-belief trials. The maximum correct false-belief score for the thought-bubbles task was 2.

False belief: Object hiding. Participants had to work out and point to which of two identical boxes contained an object. Children were told that the communicator (seated behind the hider) would help them find where an object (a balloon) was hidden by pointing to the correct box. Following Call and Tomasello (1999), three control trials were first administered. In the visible displacement trial, the hider concealed the balloon (carried out behind a raised barrier) and presented the two boxes to the child. The communicator then pointed to the box the balloon was in. Then the communicator turned around to have a drink of water and searched noisily in her bag. This distraction was natural while being obvious enough to ensure that the child knew the communicator was distracted and not paying attention to what would be happening next. In full view of the child, the hider opened up both boxes, transferred the balloon from one to the other, and closed the lids to the boxes. Then the communicator (timing was well practiced) turned around, and the hider asked the child to point to where the balloon was. This control trial confirmed that all participants could successfully keep track of the hider taking the object out from one box and moving it into the other. The next control trial was nearly identical to the first except that the hider simply switched the location of the boxes in full view of the child without taking the balloon out. The hider then asked the child to point to where the sticker was. This second control trial confirmed that all participants were able to successfully keep track of the object when the hider changed the location of the box in which the object was hidden. In the final control trial, after the hider hid the balloon and presented the boxes to the child, the communicator before pointing to the location of the object turned around to drink water and search in her bag. At this point, the hider opened both boxes and switched the balloon from one container to another and closed the lids to the containers, all in full view of the child. Then the communicator turned around and pointed to the (incorrect) container she had seen the balloon hidden. Then the hider asked the child to point to where the balloon was. This last control trial confirmed that all participants were able to successfully ignore the communicator's signal when it was wrong.

In the verbal false-belief trials, children watched the communicator observing the hider conceal the balloon. Then the hider presented the two boxes toward the child. Before the communicator pointed to which box the balloon was in, she turned around to drink and search in her bag. Then, in full view of the child, the hider opened the lid of the loaded box and moved the balloon into the other container and put the lids back on. After the hider switched the balloon, she asked the critical verbal false-belief question, "When [communicator] turns back around, where will she think the balloon is?" After the child pointed, the communicator turned back around and pointed to the box where she thought the balloon was. Two of such trials were administered and the scores were averaged.

In the two nonverbal false-belief trials, the hider hid the balloon in one of the boxes behind the barrier and presented them to the child. The communicator turned around to drink and search in her bag. The hider then swapped the location of the boxes in full view of the child. Then the communicator faced forward to signal which location she had seen the balloon hidden; thus, the box that she pointed to was incorrect, as she did not see the hider swapping the boxes. The child, who had no knowledge of the balloon's whereabouts until that moment was then asked to point to where the balloon was. Call and Tomasello (1999) argued that this trial represented a nonverbal measure of false belief because participants need to interpret the nonverbal information provided by the communicator in an appropriate way: (a) she initially saw the balloon being hidden in one of the boxes, but (b) did not see the boxes swapping locations, and (c) must have marked the particular box upon turning back around based on her outdated false belief, and so (d) the balloon must be in the other box. Two nonverbal trials were given and scores averaged. Presentation of the verbal and nonverbal false-belief blocks was counterbalanced. Average scores on the verbal and nonverbal false-belief blocks were recombined to produce a full object-hiding task score (maximum of 2).

Results and Discussion

Eye gaze and explicit false-belief performance. Replicating Study 1, a majority of the children displayed correct first looks (N = 32/36) than expected by chance ($p < .001$, binomial test). There was no significant age difference for the amount of time spent looking at the correct or the incorrect location (see Table 1 for mean values). However, children still spent a longer total amount of time looking at the correct location (M = 1.46 s) than the incorrect location (M = 0.64), $F(1, 35) = 16.62$, $p < .001$, $\eta_p^2 = .32$. As in Study 1, contingencies between passing/failing visual orienting versus giving a correct/incorrect verbal answer to the single explicit false-belief question found in Ruffman et al.'s (2001) predictive looking task were examined. There was a significant relation between eye gaze and verbal responding on the direct false-belief question, $\chi^2(1, N = 36) = 4.98$, $p < .05$, Cramer's $\phi = .37$. Sixteen children (44%) passed eye gaze and correctly answered the single explicit false-belief question while 7 children failed both (20%). However, many more children passed eye gaze while erring on the verbal answer to the explicit false-belief question (33%, $n = 12$) but only 1 participant (3%) showed the reverse relation.

Ruffman (2004) has argued that the object-hiding task is equivalent to standard false-belief measures tapping into explicit knowledge in that children are directly asked a question about the central protagonist's belief. Fitting with Ruffman's argument, many children passed eye gaze while erring on all of the false-belief test questions on the full object-hiding task (53%; $n = 19$) but only 1 participant (3%) showed the reverse relation of passing the object-hiding task but failing eye gaze. Similarly, more children passed eye gaze while erring on all of the false-belief questions on the thought-bubbles task (31%; $n = 11$) but only 2 participants (6%) showed the reverse relation of passing thought bubbles but failing eye gaze. In short, the dissociation between accurately gazing in anticipation of a character's mistaken belief while erring on verbal responses to the direct false-belief question in the predictive looking task was also found when contingencies between anticipatory gaze and overt pointing responses given for the object-hiding and thought-bubbles tasks were considered. Furthermore, not only were the verbal and nonverbal subparts of object-hiding task correlated ($r = .71$, $p < .001$), but children's scores on the overt false-belief questions asked for unexpected contents, deceptive appearance, change in location, thought bubbles, and verbal and nonverbal parts of the object-hiding task were intercorrelated ($\alpha = .87$).

In relation to bivariate associations with complex language ability, children's scores to overt false-belief questions for the change-in-location, deceptive appearance, and unexpected-contents task contexts were each correlated with complement mastery ($r = .48$, $p < .01$; $r = .54$, $p < .01$; and $r = .36$, $p < .05$, respectively). Children's tallies on the thought-bubbles task alongside scores on the verbal and nonverbal parts of the object-hiding task were also correlated with complement mastery ($r = .49$, $p < .01$; $r = .40$, $p < .05$; $r = .38$. $p < .05$, respectively). These results replicate work with deaf children (e.g., Schick et al., 2007) and could be claimed to fit with the view that complement mastery is important for any false belief reasoning. However, these particular false-belief tasks tap explicit knowledge by posing direct questions to children; as such, it is not surprising that they should all correlate positively with complement mastery (Ruffman, 2000, 2004). Critically, and anticipating the rest of the results, eye gaze remained separate by being *not* correlated with complement mastery. Consequently, a composite *explicit* false-belief understanding score was calculated by averaging subscores on answers to direct questions for unexpected contents, change in location, deceptive appearance, thought bubbles, and the full object-hiding task (hence, average maximum composite total = 2). Four-year-olds obtained significantly higher composite explicit false-belief scores compared to 3-year-olds (see Table 1).

Correlations and hierarchical regressions. Mean scores on the remaining measures of interest are also found in Table 1, and the direction of age group differences replicate those uncovered in Study 1. The analyses then turned to investigate correlations between anticipatory eye gaze (measured by the difference metric, as per Study 1) and composite explicit false-belief understanding with other variables of interest. Bivariate and partial correlations are reported in Table 2 (second row, above diagonal). At the bivariate level, PPVT performance, complement mastery, cognitive flexibility, and explicit false-belief understanding were related to chronological age. PPVT and complementation scores were also interrelated, and explicit false-belief reasoning was correlated with both PPVT and complementation performance. Explicit false-belief reasoning was further correlated with anticipatory eye gaze and cognitive flexibility. General nonverbal ability on the Matrices was also correlated with anticipatory gaze. Complement mastery was also correlated with cognitive flexibility at the bivariate level. Upon partial correlation analysis, only the following results remained: (a) explicit false-belief reasoning was related to anticipatory eye gaze performance, (b) explicit false-belief reasoning was related to cognitive flexibility and complement mastery, and (c) complement mastery was correlated with cognitive flexibility.

Two hierarchical regression analyses were conducted to determine whether implicit knowledge reflected in anticipatory eye gaze, together with cognitive flexibility and complementation, contributed unique variance to the prediction of explicit false-belief understanding. Age, and PPVT and Raven's scores were entered as the first block of independent variables. In the first regression analysis, eye gaze performance and cognitive flexibility were entered in the second and third steps, respectively. The second regression analysis was similar to the first except that complement mastery was entered at the final step. I did not pit cognitive flexibility against complement mastery when considering their unique effects because they were correlated, and it is possible that complex language might service the work of cognitive flexibility in promoting explicit false-belief reasoning (e.g., Müller et al., 2005; a point taken up in the General Discussion). Regressions are summarized in Table 4.

In the first regression analysis, age was the only significant predictor at Step 1, and accounted for 34% of explicit false-belief variance. Implicit (eye gaze) knowledge and cognitive flexibility contributed uniquely to explicit false-belief variance after effects associated with age, PPVT performance, and nonverbal reasoning on the Matrices were jointly removed. They accounted for a further 13% and 10% of the variance, respectively. For the second regression analysis, complement mastery significantly accounted for a further 9% of the variance after the effects of implicit knowledge was removed. Overall, implicit knowledge, cognitive flexibility and complement mastery contributed unique variance to predicting explicit false-belief understanding.

Study 3

The predictive looking task set up in Studies 1 and 2 involved live experimenters fiddling about

Table 4

Results of Hierarchical Regressions for Variables Predicting Explicit False-Belief Performance in Study 2 (N = 36)

Variable	B	$SE\ B$	β	T	ΔR^2	ΔF	Sign. ΔF
Regression 1							
Step 1							
Age	.06	.02	.48	2.91**	.34	5.51	.00
Matrices	−.01	.03	−.07	−0.45			
PPVT	.01	.01	.21	1.24			
Step 2							
Age	.05	.02	.41	2.74*	.13	7.86	.00
Matrices	−.04	.03	−.19	−1.32			
PPVT	.01	.01	.18	1.21			
Eye gaze	.24	.09	.40	2.80**			
Step 3							
Age	.03	.02	.22	1.43	.10	7.24	.01
Matrices	−.03	.03	−.16	−1.16			
PPVT	.01	.01	.23	1.66			
Eye gaze	.25	.08	.42	3.18**			
Cog. flexibility	.33	.12	.36	2.69*			
Regression 2							
Step 3							
Age	.04	.02	.32	2.22*	.09	6.56	.01
Matrices	−.04	.03	−.22	−1.61			
PPVT	.01	.01	.13	0.90			
Eye gaze	.21	.08	.34	2.52*			
Complements	.19	.07	.35	2.56*			

Note. PPVT = Peabody Picture Vocabulary Test.
$^{*}p < .05$. $^{**}p < .01$.

with the cardboard house and dolls in front of the child. It is possible that the timing and accentuation of events was not as rigorously controlled as it would be with a fixed video. It is important to re-examine anticipatory gaze while ensuring that all participants witnessed the unfolding of the episodes in the predictive looking task in a time-controlled fashion. Moreover, Study 2 was the only one that included the DCCS measure; it is important to replicate the findings that cognitive flexibility, alongside complement mastery, contributed unique variance to explicit false-belief understanding. A third study was carried out where Ruffman et al.'s (2001) predictive looking task was staged and filmed first and only then played back on a monitor to all participants. This change ensured that the timing and accentuation of events to elicit eye gaze was rigorously controlled for all participants. This study again included standard measures of explicit false-belief understanding together with the complement mastery and DCCS tasks. It was predicted that eye gaze would not be correlated with complex language and that cognitive flexibility and complement mastery would contribute unique variance to the prediction of explicit false-belief reasoning.

Method

Participants. New data came from 42 participants: twenty-one 3-year-olds ($M = 3.4$, SD = 3.78 months, range = 3.0–3.10; 13 girls and 8 boys) and twenty-one 4-year-olds ($M = 4.5$, SD = 2.79 months, range = 4.1–4.10; 9 girls and 12 boys). A total of 46 children were tested but three 3-year-olds and one 4-year old were not included in the final analysis as they failed to answer control questions in the false belief tasks correctly.

Procedure. Participants were individually tested in a quiet room at their kindergarten, and all took part in four 10- to 15-min sessions spaced about 1 week apart from each other. Tasks were in the same order as in Studies 1 and 2. Except for two methodological changes, Study 3 replicated the tasks used in Study 2.

First, with respect to the predictive looking task, the only difference was that the episodes in the gaze task (for false- and true-belief trials) were enacted and filmed, and the digitized film clips with accompanying audio track were then played back to children on a 15.4-in. (diagonal) monitor. The film clips were digitized into mpeg formats such that most of the screen was filled upon playback—this ensured that the correct and incorrect locations in the false- and true-belief stories were mapped toward either ends of the monitor. At least to a certain extent, then, the linear conditions of observing the screen was maximized to ensure that the whites of participants' eyes would produce strong contrast with the pupils upon looking left or right. In this way, steps were taken to ensure that anticipatory eye gaze patterns were, as much as possible, straightforward to interpret and analyze. Each participant sat in front of the computer monitor (on a table) and a camera was placed above it to record eye gaze. One rater coded all recordings of the false-belief trial in slow motion and scored gaze direction on a frame-by-frame basis (all participants passed the true-belief control predictive looking trial). Another rater coded 20% of the videotapes, and raters agreed 100% of the time as to whether or not the child was looking more to the correct location.

Second, for the explicit false-belief battery, the thought-bubbles and object-hiding tasks were not used; Study 2 showed that those tasks mapped onto standard measures of direct false-belief reasoning. Following Study 1, then, only single trials of the standard M&Ms, sponge-rock, and Sally-Ann false-belief tasks were administered. Correct verbal answers on the direct test questions based on the contexts of unexpected contents, deceptive appearance, and change in location each contributed a subtotal of 2 points to explicit false-belief knowledge; this yielded a combined maximum score of 6 for each participant ($\alpha = .70$).

Results and Discussion

Eye gaze. Replicating Study 2, significantly more children showed correct first looks ($N = 37/42$) than expected by chance based on the binomial distribution ($p < .001$). There were no significant age group differences for the amount of time spent looking at the correct or the incorrect location (see Table 1). As in Studies 1 and 2, children still spent a longer total amount of time looking at the correct location ($M = 1.87$ s) than the incorrect location ($M = 0.63$), $F(1, 41) = 60.14$, $p < .001$, $\eta_p^2 = .60$. There was also a significant link between accurately gazing in anticipation of a character's mistaken belief and giving a correct verbal answer for the explicit false-belief question, $\chi^2(1, N = 42) = 4.68$, $p < .05$, Cramer's $\phi = .33$. Nineteen children (45%) passed eye gaze and correctly answered the single explicit false-belief question while 5 children failed both (12%). However, 43% ($n = 18$) of children passed eye gaze while erring on their verbal responses to the explicit false-belief question, but no participant showed the reverse relation. Beyond eye gaze, group differences on the remaining measures of interest also replicated the findings of Study 2 (see Table 1).

Correlations and hierarchical regressions. Full bivariate and partial correlations are reported in Table 2 (second row, below diagonal). As in Study 2, when the effects of chronological age and general developmental abilities were controlled, explicit false-belief reasoning was related to eye gaze, cognitive flexibility, and complement mastery. Anticipatory eye gaze continued to show no significant correlation with complement mastery. Cognitive flexibility and complement mastery remained correlated during full partial correlation analysis. Following analyses in Study 2, two hierarchical regressions were conducted to clarify whether implicit (gaze) knowledge, alongside cognitive flexibility and complex syntactical understanding contributed unique variance to the prediction of explicit false-belief reasoning (see Table 5).

In the first regression analysis, PPVT performance was a significant predictor at Step 1, and accounted for 30% of explicit false-belief variance. Implicit knowledge and cognitive flexibility contributed uniquely to explicit false-belief variance after effects associated with age, PPVT performance, and nonverbal reasoning on the Matrices were jointly removed. They accounted for a further 18% and 7% of the variance, respectively. Interestingly, in Step 3 when all variables were considered in the first regression, PPVT performance no longer contributed unique variance to predicting explicit false-belief scores. General lexical knowledge via the PPVT may have predicted explicit false-belief performance at Step 1 because, on some superficial level, those verbal theory-of-mind tasks require some level of language ability (e.g., Chandler, Fritz, & Hala, 1989). However, the PPVT taps not only into word knowledge but also discourse semantics (e.g., words have a labeling function). Jacques and Zelazo (2005) explained that labeling can also help theory-of-mind performance by facilitating cognitive flexibility, which in turn promotes higher level consciousness in thought and control over explicit false-belief attribution. They also posited that

Table 5
Results of Hierarchical Regressions for Variables Predicting Explicit False-Belief Performance in Study 3 (N = 42)

Variable	*B*	*SE B*	β	*T*	ΔR^2	ΔF	Sign. ΔF
Regression 1							
Step 1							
Age	.06	.04	.23	1.39	.30	5.38	.00
Matrices	−.13	.10	−.19	−1.35			
PPVT	.04	.02	.39	2.31*			
Step 2							
Age	.05	.04	.19	1.28	.18	12.62	.00
Matrices	−.08	.09	−.12	−0.95			
PPVT	.03	.02	.31	2.07*			
Eye gaze	.81	.23	.45	3.55**			
Step 3							
Age	.01	.04	.04	0.25	.07	5.36	.02
Matrices	−.03	.09	−.04	−0.32			
PPVT	.03	.02	.28	1.97			
Eye gaze	.75	.22	.41	3.46**			
Cog. flexibility	.68	.29	.32	2.31*			
Regression 2							
Step 3							
Age	.05	.03	.19	1.41	.12	10.18	.00
Matrices	−.04	.08	−.05	−0.46			
PPVT	.01	.02	.07	0.46			
Eye gaze	.81	.20	.44	3.99**			
Complements	.60	.19	.41	3.19**			

Note. PPVT = Peabody Picture Vocabulary Test.
$^*p < .05$. $^{**}p < .01$.

understanding the arbitrary and symbolic nature of labels themselves could allow for psychologic distancing, which in turn promotes cognitive flexibility. Either way, the effects of semantic ability (lexical and discourse semantics) via the PPVT for explicit false-belief performance may be partly because of effects on cognitive flexibility. This may explain why PPVT scores were no longer significant once cognitive flexibility performance was entered at the final step of the regression. At the final step of the second regression analysis, complement mastery significantly accounted for 12% of the variance after the effects of implicit knowledge was removed. Even when all variables were considered, implicit knowledge continued to uniquely predict explicit false-belief understanding.

General Discussion

According to de Villiers (2000, 2004, 2005), the structural properties of tensed complement syntax provide the critical breakthrough format that allows false-belief understanding. Embedded within this view are the related claims that children's nonverbal and verbal false-belief task performance pattern after each other in their dependency upon complement mastery and that children are not likely to demonstrate any false-belief reasoning that is fugitive from linguistic propositional representation. Fitting with previous research (e.g., Milligan et al., 2007), Studies 1–3 revealed a significant relation between complement mastery and false-belief understanding. However, the current findings go beyond extant research by revealing that the relation is only evident for tasks that directly ask children to make false-belief decisions and not for tasks that indirectly elicit false-belief understanding. These findings elevate theorizing on language and theory of mind by weaving in data on children's implicit and explicit understanding.

Implicit understanding of false belief was confirmed across various analyses of eye gaze: correct first look, longer total looking to the correct location, and in terms of the richer metric of difference in looking time when noise because of fixations at the incorrect location was removed. Despite the relatively course-grain analyses of eye gaze, the findings follow suit with an accumulating number of studies showing that looking behavior can provide important information about distinctions between implicit and explicit knowledge (e.g., Clements & Perner, 1994; Onishi & Baillargeon, 2005; Ruffman et al., 2001; Southgate et al., 2007; Surian et al., 2007). The use of formal automatic eye-tracking technology to measure gaze fixation on well-defined points could expand the scope of analysis on the grade of implicit knowledge being tapped as a function of exact ocular spatiotemporal trajectories. Nonetheless, the present investigation provides important evidence that anticipatory looking is implicit in a new sense: Complement mastery is not correlated with gaze performance, and complex syntactical understanding is only correlated with direct false-belief performance. These results were replicated across three studies with larger sample sizes and when the predictive looking task was presented as a film to remove timing confounds. The dissociation signature whereby grasp of complementation structure relationally impacts on one form of knowledge but not another is evidence for separate implicit and explicit representational systems.

In Study 2 where other less verbal false-belief tasks such as object hiding and thought bubbles were included, the pattern of findings replicated Schick et al.'s (2007) work showing a significant relation between responding on these kinds of tasks with complement mastery. While anticipatory gaze performance was correlated with object-hiding and thought-bubbles false-belief scores, eye gaze on the predictive looking task still did not correlate with complement mastery. Performance on those low-verbal false belief tasks was, however, strongly correlated with verbal answers provided for traditional false-belief tasks (i.e., change in location, unexpected contents, and deceptive appearance). It seems that nonverbal false-belief attribution patterns after direct or verbal answers only insofar as they tap into explicit theory of mind understanding (Ruffman, 2000, 2004).

In the present investigation, there was no significant difference between 3- and 4-year-olds' implicit false-belief knowledge as tapped by the predictive looking task. The issue of whether implicit knowledge, despite its earlier manifestation, should show ontogenetic maturation is a gnawing one. If the present investigation had studied a wider age range (e.g., included 2-year-olds), there may be significant group differences in the extent of implicit knowledge as tapped by the usual predictive looking task, and the difference in looking time variable may even then correlate with chronological age (e.g., Clements & Perner, 1994; Ruffman, 2000). Gradual change in implicit knowledge may also be suggested when we consider children's performance on a variety of indirect false-belief tasks. For example, 15-month-olds show implicit false-belief sensitivity via the violation-of-expectancy methodology (e.g., Onishi & Baillargeon, 2005), and 2-year-olds show implicit false-belief sensitivity when the predictive looking task is simplified by completely removing the target object from the scene (e.g., Southgate et al., 2007). It is possible that children may be building up implicit representations of false belief with a greater number of processing units and stronger connection weights, as embodied in a neural network, and the different types of indirect false-belief tasks may necessitate different grades of implicit network representation.

Mechanismwise, implicit knowledge could develop in part through statistical pattern learning. Children may detect regularities across events where people often successfully locate objects where they last put them. Across most events, these regularities work well, but occasionally children may be confronted by actions where people look for objects in the wrong place. Other inferential systems such as goal or intentional tracking could potentially help parse these actions and contexts so that an implicit false-belief interpretation is discerned over event irregularities where people end up going to look for an object not where it necessarily is. Also, because such events are less common compared to people successfully looking for objects where they currently are, it takes time for implicit false-belief knowledge to be built up with sufficient network complexity (Clements et al., 2000). Consequently, there will be phases in knowledge construction whereby young children show fragile anticipatory gaze, and also phases where young children may show dissociation in false-belief understanding at the levels of spontaneous visual orienting and verbal judgment. Further detections of exceptions to behavioral regularity may result in a state of cognitive disequilibrium that critically pushes systemwide integration of all extant knowledge representations into a higher order format that supports conscious and verbally correct false-belief judgments. Such a theoretic possibility could potentially explain why, as found in Studies 2 and 3, explicit false-belief understanding also correlated significantly with flexibility in handling higher order rules that is partly tapped by the DCCS. Of course, the current data on early implicit false-belief understanding in 3- and 4-year-olds cannot decisively arbitrate on whether the basis of such rudimentary knowledge should be viewed as constructivist and hard-won or nativist in origin. One could instead argue that the high accuracy in first look reflects a modular theory-of-mind mechanism (e.g., Leslie et al., 2005). Leslie et al. (2005) also view the theory-of-mind mechanism to be assisted by its own on-board selection processor that inhibits true-belief suggestions. As young children have limited inhibitory control, they may automatically self-correct in looking between the correct (false belief) and incorrect (true belief) locations, leading to variation over the difference in looking time metric, as was found here. To the extent that an early rudimentary understanding of false belief might be supplied by a modular mechanism, and basic recognition is progressively made explicit through intersections with higher order systems of language and aspects of executive control, there is overlap between constructivist and nativist discussions of the current findings.

Reiterated rounds of internal redescription of implicit knowledge yielding stronger representations of explicit knowledge may come about when children engage in cognitive flexibility over the course of problem solving (Karmiloff-Smith, 1992; Zelazo, 1994). Replicating previous work (e.g., Frye et al., 1995; Müller et al., 2005), Studies 2 and 3 revealed that DCCS performance was positively correlated with composite explicit false-belief performance. This association could have partly come about through cognitive flexibility in constructing an explicit higher order conditional if-if-then rule helping to embed different perspective settings so that conflicting knowledge can be accurately applied. For instance, in the object-hiding direct false-belief task, embedded conditionals might be: if friend turned around (Setting 1), and if she indicated to the object being here, then predict that object is there; but if friend kept observing (Setting 2), and if she indicated to the object being here, then predict that object is here. The result that complement mastery also predicted explicit reasoning could be interpreted to dovetail with Zelazo and colleagues' broader view that complex language may potentially off-load cognitive flexibility for explicit false-belief reasoning (e.g., Jacques & Zelazo, 2005; Müller et al., 2005). It is possible that cognitive flexibility over explicit false-belief reasoning may be more effective if rules are carried out in linguistic formulations; discursive thought may create richer psychologic distance, thereby helping children explicitly understand that perspectives (and rules) are arbitrary or necessarily representational. Rule talking one's way through false-belief problems could also bring out explicit thinking indirectly via the effect of linguistic formulations making psychologic perspectives an object of consideration.

However, we need to carefully consider the implication that mastery of complement syntax per se also predicted children's explicit false-belief reasoning scores. Indeed, in interpreting the correlation between DCCS performance and complement mastery found in Studies 2 and 3, it is difficult to think of how preschoolers could have solved the memory for complements task by necessarily using if-if-then propositional language. Proficiency in hierarchical rule use, as partly tapped by the DCCS task, could provide for accurate conscious false-belief problem solving, but only within limits. Preschoolers could also explicitly compare two representations of a situation through syntactic constructions involving a single verb of belief and an embedded tensed complement (e.g., "Sally thinks *the ball is in the basket*"). In the same way, for the object-hiding task in Study 2, preschoolers could explicitly recognize perspectives and make an accurate judgment when they verbally formulate the communicator's signal as "the friend says (indicates)/thinks *the object is there.*" Understanding richness of minds may also be partly explicated using communication and mental state verb-complement constructions. Nonetheless, there is still the question of how precisely cognitive flexibility and complement mastery might come to developmentally interface or intersect to underpin explicit false-belief understanding. One modest interpretation of the current results may simply be that linguistically mediated thought involving diverse syntactic constructions at different levels of analysis (e.g., from pruned complements to if-if-then rules) is partly responsible for conscious flexibility in shifting from one response perspective to another (e.g., Carlson, 2003; Smith, Apperly, & White, 2003).

The empirical finding that complement mastery was uniquely related to explicit false-belief understanding raises a further thorny issue: Even though participants were not invited to consider the motives of the false utterance, the misstatements in the complements tasks potentially involve misrepresentation of an actual state of affairs. Complement mastery and explicit false-belief reasoning may be correlated because they both involve misrepresentation (as opposed to a more specific role for complex syntax per se). The complements tasks could even be viewed as a simplified (but explicit) measure of false belief. It will be crucial to develop other viable analogs of the complement mastery task to delineate the extent to which it is the implications of certain syntactic constructions that partly lead children to explicitly understand differences in minds. The development of implicit measures (e.g., capitalizing on looking time) could open up investigations as to whether young infants and preschoolers may show nascent sensitivity to how a false proposition can be legally attached to a verb of communication. In doing so, we stand to discover whether an implicit understanding of complement syntax may be related to an implicit understanding of false belief. Finally, we need to test children on multiple measures of indirect theory-of-mind tasks that also include controls for confounds because of associative encoding (e.g., Surian et al., 2007)—higher reliability from composite scores of implicit false belief may uncover whether there are weak, but significant, correlations between rudimentary false-belief knowledge and complement mastery. Overall, it will be important for future research to discover how early implicit understandings that intertwine the beginnings of theory of mind and language acquisition metamorphosize into more explicit knowledge (Astington, 2005).

In conclusion, a richer understanding of children's theory of mind will require collaborative efforts that use multiple methods, theories, and levels of analysis to illuminate how children's conceptual representations may be differentiated and integrated along the implicit–explicit continuum. Explicit false-belief understanding was directly underpinned by unique implicit false-belief knowledge as tapped by anticipatory eye gaze. This finding is important because it means that the causal role of complex language in the development of false-belief reasoning does have to be constrained, to refer not to any reasoning but only to explicit answering. Indeed, aside from preschoolers being able to operate on the basis of automatic implicit knowledge, explicit verbal false-belief responding was also multiply supported by complement mastery and cognitive flexibility. Generally, to the extent that explicit false-belief understanding is open to input from language and higher order systems of executive control, the current findings on distinctions between implicit and explicit false-belief reasoning have significant empirical and theoretic fit with different levels of analysis interested in how growth in reflection makes possible significant cognitive advances. While the answers uncovered in this investigation turned out to be fairly complex, integrative work of the kind attempted here serve as an important start to answering questions on how representational changes in mental state understanding develop as a function of intersections between language, cognitive control, and early core knowledge.

References

Astington, J. W. (2005). The developmental interdependence of theory of mind and language. In N. J. Enfield & S. C. Levinson (Eds.), *Roots of human sociality: Culture, cognition and interaction* (pp. 179–206). Oxford, UK: Berg.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition, 21*, 37–46.

Bialystok, E., & Senman, L. (2004). Executive processes in appearance-reality tasks: The role of inhibition of attention and symbolic representation. *Child Development, 75*, 562–579.

Bornstein, M. H., & Sigman, M. D. (1986). Continuity in mental development from infancy. *Child Development, 57*, 251–274.

Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science, 8*, 535–543.

Call, J., & Tomasello, M. (1999). A non-verbal false-belief task: The performance of children and great apes. *Child Development, 70*, 381–395.

Carlson, S. M. (2003). Executive function in context: Development, measurement, theory and experience. *Monographs of the Society for Research in Child Development, 68*(Serial No. 274).

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72*, 1032–1053.

Carlson, S. M., Moses, L. J., & Claxton, L. J. (2004). Individual difference in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology, 87*, 299–319.

Chandler, M., Fritz, A. S., & Hala, S. (1989). Small-scale deceit: Deception as a marker of two, three-, and four-year-olds' early theories of mind. *Child Development, 60*, 1263–1277.

Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development, 9*, 377–395.

Clements, W. A., Rustin, C. L., & McCallum, S. (2000). Promoting the transition from implicit to explicit understanding: A training study of false-belief. *Developmental Science, 3*, 81–92.

de Villiers, J. (2000). Language and theory of mind: What are the developmental relationships? In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), *Understanding other minds: Perspectives from developmental cognitive neuroscience* (2nd ed., pp. 83–123). New York: Oxford University Press.

de Villiers, J. G. (2004). *Why language first? Invited paper, Web conference on co-evolution of language and theory of mind.* Retrieved August 19, 2008, from http://www.interdisciplines.org/coevolution/papers/5

de Villiers, J. G. (2005). Can language acquisition give children a point of view? In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 186–219). New York: Oxford University Press.

de Villiers, J. G., & de Villiers, P. A. (2000). Linguistic determinism and the understanding of false-beliefs. In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning and the mind* (pp. 189–226). Hove, UK: Psychology Press.

de Villiers, J. G., & de Villiers, P. A. (2003). Language for thought: Coming to understand false-beliefs. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 335–384). Cambridge: MIT Press.

de Villiers, J. G., & Pyers, J. E. (2002). Complements to cognition: A longitudinal study of the relationship between complex syntax and false-belief understanding. *Cognitive Development, 17*, 1037–1060.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test (PPVT–III)* (3rd ed.) Circle Pines, MN: American Guidance Service.

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule based reasoning. *Cognitive Development, 19*, 483–528.

Garnham, W. A., & Ruffman, T. (2001). Doesn't see, doesn't know: Is anticipatory looking really related to understanding of belief? *Developmental Science, 4*, 94–100.

Goldin-Meadow, S., & Alibali, M. W. (1999). Does the hand reflect implicit knowledge? Yes and no. *Behavioral and Brain Sciences, 22*, 766–767.

Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false-belief and the appearance-reality distinction. *Child Development, 59*, 26–37.

Hale, C. M., & Tager-Flusberg, H. T. (2003). Influence of language on theory of mind. *Developmental Science, 6*, 346–359.

Hongwanishkul, D., Happaney, K. R., Lee, W. S. C., & Zelazo, P. D. (2005). Assessment of hot and cool executive function in young children: Age-related changes and individual differences. *Developmental Neuropsychology, 28*, 617–644.

Jacques, S., & Zelazo, P. D. (2005). Language and the development of cognitive flexibility: Implications for theory of mind. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 144–162). New York: Oxford University Press.

Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science.* Cambridge, MA: MIT Press.

Kirkham, N. Z., Cruess, L., & Diamond, A. (2003). Helping children apply their knowledge to their behavior on a dimension-switching task. *Developmental Science, 6*, 449–476.

Kloo, D., & Perner, J. (2003). Training transfer between card sorting and false belief understanding: Helping children apply conflicting descriptions. *Child Development, 74*, 1823–1839.

Leslie, A. M., German, T. P., & Pollizi, P. (2005). Belief-desire reasoning as a process of selection. *Cognitive Psychology, 50*, 45–85.

Lohmann, H., & Tomasello, M. (2003). The role of language on the development of false-belief understanding: A training study. *Child Development, 74*, 1130–1144.

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development, 78*, 622–646.

Miyuki, K., & Chisato, Y. (2006). Understanding gaze perception and control. *Japanese Journal of Psychonomic Science, 25*, 77–84.

Müller, U., Zelazo, P. D., & lmrisek, S. (2005). Executive function and children's understanding of false belief: How specific is the relation? *Cognitive Development, 20*, 173–189.

Munakata, Y., McClelland, J. L., Johnson, M. H., & Siegler, R. S. (1997). Rethinking infant knowledge: Toward an adaptive process account of successes and failures in object permanence tasks. *Psychological Review, 104*, 686–713.

O'Neill, D. K. (1996). Two-year-old children's sensitivity to parent's knowledge state when making requests. *Child Development, 67*, 659–677.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*, 255–258.

Perner, J., & Clements, W. A. (2000). From an implicit to an explicit theory of mind. ln Y. Rossetti & A. Revonsuo (Eds.), *Beyond dissociation: Interaction between dissociated implicit and explicit processing* (pp. 273–294). Amsterdam: John Benjamins.

Perner, J., & Lang, B. (2000). Theory of mind and executive function: ls there a developmental relationship? ln S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), *Understanding other minds: Perspectives from autism and developmental cognitive neuroscience* (pp. 150–181). New York: Oxford University Press.

Perner, J., Stummer, S., Sprung, M., & Doherty, M. (2002). Theory of mind finds its Piagetian perspective: Why alternative naming comes with understanding belief. *Cognitive Development, 17*, 1451–1472.

Raven, J. C. (1993). *Colored Progressive Matrices*. Oxford, UK: Oxford Psychologists Press.

Ruffman, T. (2000). Nonverbal theory of mind: ls it important, is it implicit, is it simulation, is it relevant to autism? In J. W. Astington (Ed.), *Minds in the making: Essays in honor of David R. Olson* (pp. 456–479). Oxford, UK: Blackwell.

Ruffman, T. (2004). *Theory of mind and language ability: Understanding the bigger picture. Invited paper, Web conference on co-evolution of language and theory of mind.* Retrieved August 19, 2008, from http://www.interdisciplines.org/coevolution/papers/8

Ruffman, T., Garnham, W., Import, A., & Connolly, D. (2001). Does eye gaze indicate implicit knowledge of false-belief? Charting transitions in knowledge. *Journal of Experimental Child Psychology, 80*, 201–224.

Schick, B., de Villiers, P., de Villiers, J., & Hoffmeister, R. (2007). Language and theory of mind: A study of deaf children. *Child Development, 78*, 376–396.

Smith, M., Apperly, l. A., & White, V. (2003). False belief reasoning and the acquisition of relative clause sentences. *Child Development, 74*, 1709–1719.

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false beliefs by two-year-olds. *Psychological Science, 18*, 587–592.

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-olds. *Psychological Science, 18*, 580–586.

Wellman, H. M., Phillips, A. T., Dunphy-Lelii, S., & LaLonde, N. (2004). lnfant social attention predicts preschool social cognition. *Developmental Science, 7*, 283–288.

Woolfe, T., Want, S. C., & Siegal, M. (2002). Signposts to development: Theory of mind in deaf children. *Child Development, 73*, 768–778.

Zelazo, P. D. (1994). From the decline of development to the ascent of consciousness. *Behavioral and Brain Sciences, 17*, 731–732.

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols, 1*, 297–301.

Zelazo, P. D., Gao, H. H., & Todd, R. (2007). The development of consciousness. ln P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *Cambridge handbook of consciousness* (pp. 405–432). New York: Cambridge University Press.

Child Development, March/April 2010, Volume 81, Number 2, Pages 616–635

Bridging Theory of Mind and the Personal Domain: Children's Reasoning About Resistance to Parental Control

Kristin Hansen Lagattuta
University of California, Davis

Larry Nucci
University of California, Berkeley

Sandra Leanne Bosacki
Brock University

Four-, 5-, and 7-year-olds ($N = 60$) listened to vignettes featuring characters that wanted to do actions that conflicted with parental rules. Desires included behaviors associated with the personal domain: friend, activity, and clothing choice. Scenarios involving moral rules served as a comparison. Children predicted and explained characters' actions and emotions. Results showed significant increases between 4 and 7 in judgments that characters would comply with rules and feel good, but only for situations involving moral rules. Children frequently predicted that characters would disobey rules that intruded on the personal domain and would feel positive emotions following noncompliance, especially when activities were essential to that character's identity. Findings are discussed in relation to the development of self and personal control.

Young children frequently confront situations where they really want to do something, but their desire conflicts with parental or societal rules. The ability to inhibit personal impulses to abide by rules and norms comprises a major developmental achievement and a central component of moral development (Kochanska & Aksan, 2006; Thompson, Meyer, & McGinley, 2006). Piaget (1962) referred to the affective conflicts that arise between moral duty and desire as the ''problem of will.'' Such acts of willpower pose a difficult challenge for young children, who have less developed inhibitory control (Zelazo & Mueller, 2002) and who generally view positive emotions as resulting from desire fulfillment and negative emotions as caused by not getting what one wants (Harter & Whitesell, 1989; Stein & Levine, 1989; Wellman & Banerjee, 1991). Indeed, studies show that children younger than 7 years of age typically fail to recognize that people can experience positive emotions after inhibiting desires to abide by rules or that people can feel negative emotions after breaking rules to get what they want (Arsenio & Lover, 1995; Keller, Lourenço, Malti, & Saalback, 2003; Lagattuta, 2005; Nunner-Winkler & Sodian, 1988).

This developmental shift from 4 to 7 years in judging that rule breakers feel good to an appreciation that rule breakers can experience negative emotions is considered a significant advance in children's sociomoral reasoning, and it has been widely replicated in the United States and abroad (see Arsenio, Gold, & Adams, 2006). This transition also marks an important coordination between children's understanding of ''mental life'' or ''theory of mind'' and their reasoning about morality and societal norms (Lagattuta, 2005). That is, a core feature of everyday folk psychology is that desire motivates action: Getting what one wants feels good and not getting what one wants elicits negative emotions (see Wellman, 1990). Yet, because humans live in social groups, our actions are not only determined by individual desires, beliefs, or goals but also by the rules and norms that govern these social exchanges (Wellman & Miller, 2008). Between the

---

Portions of this research were presented at the biennial meeting of the Society for Research in Child Development, 2007, and at the biennial meeting of the Jean Piaget Society, 2007.

We thank Erica Vaughn, Sarah Ballantyne, Katherine Chi, Gloria Garcia, Lauren Jenkins, and Cortney Winkle for their assistance with this project. We thank Megan Hansen for creating the drawings that accompanied the story vignettes. We also express gratitude to the children who participated in this research.

Correspondence concerning this article should be addressed to Kristin Hansen Lagattuta, Department of Psychology and Center for Mind and Brain, University of California, Davis, 1 Shields Avenue, Davis, CA 95616. Electronic mail may be sent to khlaga@ucdavis.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8102-0016

ages of 4 and 7 children demonstrate increased ability to integrate theory of mind and sociomoral perspectives by more frequently predicting and explaining emotions not only in relation to people's current desires or goals (a theory of mind perspective) but also in relation to rule obligations and to future outcomes for self and others that may result from making particular decisions (see Lagattuta, 2005, 2008).

Importantly, however, moral development not only involves constructing concepts about rules and norms for what one should not do but also identifying actions that people *can* have personal discretion and control over, choices such as what to wear, what activities to do, and what friends to have (Nucci, 1996; Smetana, 2002). Work conducted by domain theory researchers indicates that children begin to identify a personal domain during early childhood through social messages and choices offered by caregivers (e.g., What shirt do you want to wear?), as well as by children's resistance and negotiation with parents over what constitutes their zone of privacy and prerogative (Killen & Smetana, 1999; Nucci & Smetana, 1996; Nucci & Weber, 1995). Because there may be legitimate grounds for disobedience when rules intrude on the personal domain, children may reason about behavioral and emotional responses to personal domain restrictions differently than other kinds of rules. Previous research reporting an increase between 4 and 7 years in judgments of ''feel good compliance'' and ''feel bad transgression'' have only included moral (harm to others), prudential (personal safety), or social-conventional rules (social norms); rules that children view as obligatory and under the jurisdiction of authority and not in the personal domain (see Smetana, 2005).

The distinction made within social cognitive domain theory among morality, convention, and the personal (Smetana, 2005; Turiel, 2002) begs for a more purposeful linkage with research and theory on children's theory of mind. The current study integrates theory of mind and social domain perspectives by investigating how 4- to 7-year-olds reason about people's behavior choices and emotions in situations where they want to do an activity from the personal domain (clothing, activity, or friendship choice), and a parent tells them they should not engage in that action. If children believe that personal domain restrictions can be disobeyed, then we would expect children to infrequently predict that people will comply with rules that govern the personal domain and infrequently judge that compliance with such rules makes people feel good. In explaining their judgments, children should largely provide goal-oriented explanations, focusing on the character's desires and preferences (desire psychology framework). Suggestive evidence comes from Arsenio (1988), who found that 5- to 11-year-olds attributed negative emotions to characters that had been told by teachers to share private information. On the other hand, if young children do not treat the personal domain uniquely when reasoning about how a person will act and feel, then we should replicate the significant increase between 4 and 7 years in predicting that compliance feels good and transgression feels bad found in previous studies featuring moral, prudential, and social-conventional rules.

Investigation into children's reasoning about behavioral and emotional reactions to personal domain restrictions further provides a new approach for examining children's understanding of the importance of the personal domain for self and identity. The differentiation of the personal domain from matters of morality and convention rests on the assumption that the fulfillment of personal desires is psychologically required for the establishment of ''self.'' According to Nucci (1996), the development of a positive sense of agency, self, and personal identity depends on children's ability to stake a claim to a personal area of actions (see also Erikson, 1968). Indeed, adolescents from diverse cultural and socioeconomic backgrounds explain that control over a personal domain is important for establishing and preserving an identity and for having the freedom ''to be oneself'' (Nucci, 1996, 2006; Smetana, 2005). Here, typical procedures involve interviewing older children (8 years+) and adolescents about which matters they feel are under their personal jurisdiction and why. Because these are obviously difficult concepts for young children to articulate in response to such questioning methods, little is known about whether younger children may recognize the significance of the personal domain for the promotion of self and identity.

To assess 4- to 7-year-olds' knowledge about the relation between identity concerns and the personal domain, we created a new methodology informed by developmental research on children's essentialist thinking. According to Gelman (2003), children (and adults) often believe that people, animals, or things are the way they are because of stable and unobservable properties, or essences. Moreover, children and adults expect essentialized qualities of a person to have strong inductive potential. Thus, for example, when 5- to 7-year-old children were presented with a character who really likes carrots, eats them all of the time, and is named a "carrot eater," they overwhelmingly predicted that she would continue to eat carrots in the future even if no one in her family liked carrots and even if her family tried to stop her (Gelman & Heyman, 1999). Nominalized phrases, such as "carrot eater," were particularly powerful in causing children to treat properties as immutable and central to people's identities. Given that even 7-year-olds most often predicted that "carrot eaters" ("creature believers," "guinea pig lovers," "early wakers") would go against their family to engage in actions essential to self supports our hypothesis that children will likely treat personal domain restrictions, especially those that forbid essentialized activities, as less subject to parental regulation than other rules.

In the current study, then, we varied stories featuring personal domain restrictions so that for some trials the currently desired, but forbidden activity was described as *essential* to the character's sense of identity, and for other trials it was described as *peripheral*. Essential stories described the character as doing the activity every day, feeling very good about him- or herself while doing the activity, and naming him- or herself in relation to the activity using a nominalized phrase (e.g., "Gloria the Painter"). In contrast, peripheral stories described the character as doing the activity only some of the time, feeling just OK about him- or herself while doing the activity, and not naming him- or herself in connection with the activity. If young children are sensitive to connections between the personal domain and identity, then they should predict less compliance and less *feel good compliance* when desired activities are described as essential versus peripheral to the character's sense of self. Children should also frequently explain character's behavior decisions and emotions in these personal-essential situations as motivated by aspects of their self or identity (e.g., "Because she likes to X," "Because he always plays with X").

To rule out the interpretation that children may simply respond to the essential condition in terms of the given action's high desirability for the protagonist, we added an additional story type in which the main character wants to commit a moral transgression to participate in the desired activity (e.g., Gloria wants to steal paints so that she can paint a picture). For these moral rule stories, the activity was also described as essential to the protagonist's identity. We predicted that children would less frequently judge that a person would violate a moral rule (e.g., "You should not steal other people's things") to carry out the highly desired essential goal compared to violating a parental directive that invaded the personal domain (e.g., "You should not paint"). These moral vignettes further serve to ground this study in previous research and to provide a direct comparison between reasoning about moral rules and personal domain restrictions. That is, whereas we expected that predictions of compliance and feel good compliance for the personal domain stories (especially the essential subtype) would remain low over age, these responses would significantly increase for trials featuring moral rules. We further predicted that children would more often cite moral-evaluative concerns (about behaviors being good, bad, right, or wrong) as well as potential future consequences for self or others when explaining character's decisions and emotions in moral rule situations versus scenarios involving personal domain restrictions.

Finally, because some children may never predict that a person would break parental rules, we included personal-essential trials where we explicitly told the participant that the character decided to fulfill his or her desire (e.g., Gloria decides to paint a picture even though her mom said she should not paint). We asked participants to explain why the character made that choice as well as to predict and explain his or her emotions. We added additional questions about whether the child had broken a rule, whether it was OK or not OK to have made that decision, and how the protagonist would feel if he or she never got to participate in that activity ever again. We hypothesized that all age groups would understand that rules had been broken and that the character would feel bad if he or she was forbidden from ever participating in a desired activity. We further predicted that there would be an age-related increase in judging that disobeying personal domain restrictions could be legitimate or "OK" to do. To avoid priming children to consider the legitimacy of rule violations, these explain disobedience trials appeared at the end of the session.

In summary, then, the current study conjoins theory of mind concerns with understanding the development of young children's conceptions of self and others as psychological beings with domain theory distinctions between morality and the personal domain. Children were presented with scenarios featuring a child protagonist who wants to engage in an action that violates a rule stated by a parent. Trials varied by whether the rule governed the personal (clothing, friendship, activity choice) versus moral domain (hitting, stealing), and by whether the desired activity was described as essential versus peripheral to that character's identity. We targeted 4- to 7-year-olds because previous studies have found significant changes in children's knowledge that transgression can feel bad and compliance can feel good during this age period. We hypothesized that children would treat personal domain restrictions differently because these directives violate claims to individual jurisdiction and control. In addition to examining children's predictions, we analyzed their explanations for why characters made certain decisions and experienced particular emotions. Here, attention was focused on children's provision of goal-, rule-, and future-oriented explanations as well as references to identity or moral-evaluative concerns. This multiple-assessment approach (analyzing predictions and explanations) has been found effective for assessing developmental changes in social reasoning during early childhood (see Lagattuta, 2005; Wellman & Lagattuta, 2000; Wellman & Liu, 2004).

Method

Participants

Participants included 60 children from three age groups: twenty 4-year-olds (M = 4.5 years, range = 4.0–4.92 years); twenty 5-year-olds (M = 5.41 years, range = 5.0–5.92 years); and twenty 7-year-olds (M = 7.44 years, range = 7.0–8.0 years), with equal numbers of males and females in each age group. Four- and 5-year-olds were primarily recruited from a preschool operated by an urban Midwestern university. Seven-year-olds were primarily recruited from a public elementary school from the same geographic area serving families with similar demographics. In total, the sample was 65% European American, 17% Asian American, 15% African American, and 3% Other. Although precise data on parental education and income level were not obtained, the majority of the sample came from middle-income families. Two additional children participated but were replaced due to experimenter error (one 4-year-old) and failure to complete the interview (one 5-year-old).

Design and Procedure

Participants listened to eight stories, each featuring a child protagonist who wants to do a particular action, but his or her desire conflicts with a prohibitive rule stated by his or her mother. Action preferences centered on four behaviors commonly associated with the personal domain (see Nucci & Weber, 1995): wanting to play a *sport* (baseball, soccer), wanting to engage in a *creative activity* (play with Legos, paint), wanting to wear a particular *clothing type* (sparkly clothes, superhero clothes), and wanting to play with a particular *friend*. One story in each content type featured a male central character and the other a female central character. The race and ethnic composition of the story characters matched the sample population. To aid in story comprehension, all vignettes were depicted on colorful 5 × 6 in. laminated pictures.

There were three different story types: *personal-essential*, *personal-peripheral*, and *moral-essential*. Here, the first part of the label refers to whether the prohibitive rule applied to the personal (athletic or creative activity, friendship, or clothing choice) or to the moral domain (stealing, causing physical harm). The second part of the label indicates whether the desired activity was described as essential (a fundamental component) versus peripheral (not very important) to that character's sense of self or identity. Table 1 summarizes the dimensions of each story type.

This experimental design is not a complete 2 × 2 factorial design because the moral-peripheral subtype was excluded to reduce the length of the experimental session and because this subtype was the least interesting from a theoretical standpoint. The two central contrasts of interest were between essential and peripheral versions of the personal domain restriction trials (i.e., in cases of rules infringing on the personal domain, does it matter whether the currently desired activity is essential vs. peripheral to one's sense of identity?) and between moral versus personal domain restriction versions of the essential core trials (i.e., when an activity is essential to one's sense of identity, does it matter whether the potential rule violation pertains to the moral vs. personal domain?). Note that for both personal domain restrictions and moral rules, the mother did not provide a reason for forbidding the activity. This allowed measurement of children's baseline judgments about people's behaviors and emotions in situations where parents do not provide explicit justifications for rules.

Personal-essential stories. For personal-essential stories, the desired activity was described as a fundamental core to the character's sense of identity and daily routine. Several approaches were used to convey this essential quality: The characters were

Table 1
Summary of Story Type Dimensions

Story type	Identity relevance: Essential	Identity relevance: Peripheral	Rule domain: Personal	Rule domain: Moral
	• Daily participation • Feels very good about self when engaging in the activity • Names self in relation to the activity	• Occasional participation • Feels OK about self when engaging in the activity • No naming	• You should not wear X clothes • You should not do X (activity) • You should not play with X (friend)	• You should not push or hit • You should not take other people's things
Personal-essential[a]	√		√	
Personal-peripheral		√	√	
Moral-essential	√			√

[a] As described in the Method section, personal-essential was the only story type that had two types of questioning: (a) predict-and-explain and (b) explain disobedience. The other two story types only had predict-and-explain questioning.

described as doing the activity everyday, naming themselves in relation to the activity by using a nominalized phrase (e.g., "Gloria the Painter"), and feeling very good about themselves when engaging in this activity. These components were informed by empirical findings in the essentialism literature (e.g., Gelman, 2003; Gelman & Heyman, 1999) showing that young children interpret this kind of phrasing and lexicalization as indicating immutable, stable, and central aspects of identity.

For example, as shown in Figure 1 (top left), "Gloria likes to paint pictures. She paints pictures almost every day. In fact, Gloria thinks of herself as 'Gloria the Painter.' 'Gloria the Painter' thinks she creates great pictures and that makes her feel happy about herself." The essential labels for the other stories included: "Max the Lego Builder," "Carlos the Baseball Player," "Linda the Soccer Player," "Sarah the Sparkle Princess" (girl who likes to wear sparkly clothes), "John the Superhero" (boy who likes to wear superhero clothes), "Amy the Best Friend," and "Tom the Best Friend."

For this story type, the central character was shown smiling and engaging in the activity in Card 1. Card 2 showed a four-square grid of the character smiling and participating in that activity four different times. For sport stories, Card 2 displayed the character playing that sport on four different occasions; for clothing choice, Card 2 showed the character wearing that type of clothing four different times; and for friendship choice, Card 2 depicted the character playing with the same friend on four different occasions (the friend had a distinctive hat or hair bow).

In the final scene of the personal-essential stories (see Figure 1, bottom row), the character wants to engage in the desired action, the mother states a prohibitive rule, and then the mother leaves the scene and the protagonist thinks about what to do. For example, "Well, the next day, Gloria needs to find something to do. Right now, Gloria wants to paint pictures but her mom says, 'Gloria, you should not paint pictures!' Gloria's mom leaves the room. Gloria decides what to do next." Picture Card 3 showed the main character and the mother-facing profile toward the desired activity (e.g., the easel with paper and paints). For the final scene, the mother cutout was lifted off of Card 3 so the protagonist was featured alone in the scene.

Personal-peripheral stories. Personal-peripheral stories described the character as participating in the particular activity only infrequently and as feeling just "okay" when engaged in that activity. For these stories, the character also did not name him- or herself in relation to the activity (e.g., Gloria is just referred to as "Gloria" not "Gloria the Painter"). Here, the aim was to establish that the character did not typically engage in the activity because it was not a stable, core, aspect of his or her identity.

For example, as shown in Figure 1 (top right), "Gloria likes to paint pictures some of the time. But most of the time she does other things. Gloria thinks she paints okay pictures and that makes her feel okay about herself." For this story version, Card 1 displayed the central character standing and smiling (not engaged in the desired activity) and Card 2 showed the character engaged in the targeted activity one time with a neutral expression and smiling while participating in three other activities (e.g., reading, swinging, walking). Card 2 for sport stories featured three other sport-related activities (swimming, kicking a ball, jump rope),

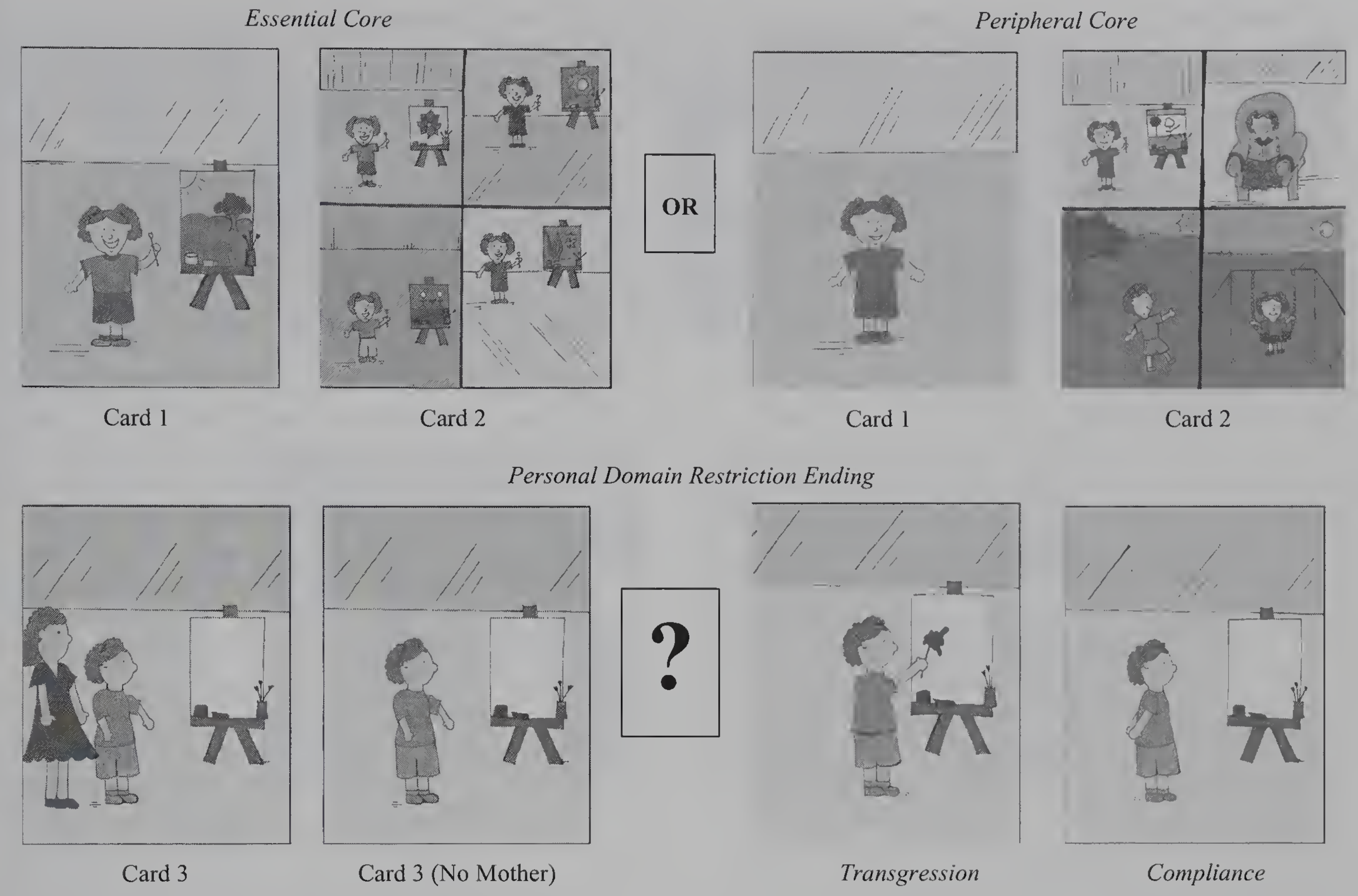


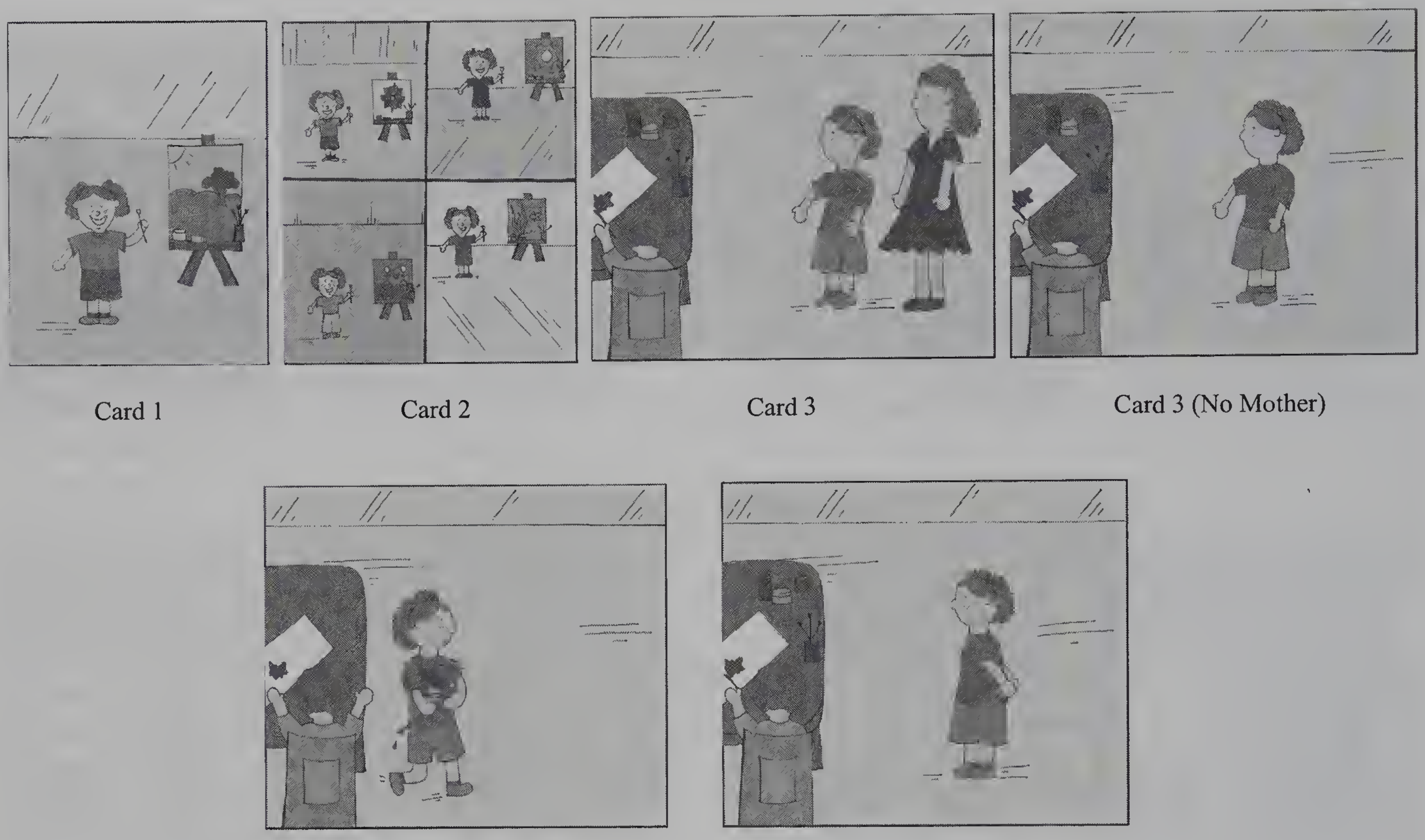
Figure 1. Example illustrations for a personal domain restriction story.

Card 2 for clothing choice stories featured the character wearing three other outfits, and Card 2 for friendship choice stories showed the protagonist playing with three other friends.

The final scene and story cards for the personal-peripheral stories were identical to that of the personal-essential stories (see Figure 1, bottom row).

Moral-essential stories. As described, moral-essential stories were created as a contrast to the personal-essential story versions. Of interest was whether young children would predict that in cases where the character identified very strongly with the activity choice he or she would even break moral rules (victimize other people) to fulfill his or her desire. Thus, moral-essential stories started with the identical essential core of the personal-essential stories. What differed was the rule type in the final scene. In the final scene, the central character wants to break a moral rule (steal or hit) to engage in the desired activity. For example as shown in Figure 2, ''Well, the next day Gloria needs to find something to do. Right now, Gloria wants to grab away her brother's paintbrush and paints so she can use them, but her mom says, 'Gloria, you should not take away your brother's paints!' Gloria's mom leaves the room. Gloria decides what to do next.'' Again, as with the personal domain restriction stories, the central character and the mother are shown in the profile in Card 3 and then the mother card was removed.

For the moral-essential stories featuring stealing as the moral rule (clothing, creative activity), the potential victim was always shown from behind in Card 3 with the desired belongings situated next to or behind him or her. For moral-essential stories featuring the desire to hit (friendship, sports), the potential victim was shown telling the main protagonist that he or she cannot play. In these cases, the protagonist wants to ''push/hit the boy/girl hard so he/she can play [soccer/baseball/with a best friend]'' and the mother states, ''You should not push/hit.'' For example, the Tom story read, ''Well, the next day at the park Tom sees another boy playing with Mark [who was previously described as Tom's best friend]. The boy says, ''Go away, Tom, you can't play!'' Right now, Tom wants to push the boy hard so he can play with Mark, but



Card 1 Card 2 Card 3 Card 3 (No Mother)

Transgression *Compliance*

Figure 2. Example illustrations for a moral-essential story.

Tom's mom says, "You should not push other people."

Questioning

Control questions. Several control questions assessed young children's ability to follow the premises of each story. Three control questions—*preference, self-evaluation*, and *self-relevance*—followed the initial core of each vignette where the experimenter described details about whether the desired activity was essential versus peripheral to that character's identity. More specifically, children were asked, "Does [character] like to [action] all of the time or only sometimes?" (preference) and "Does [character] feel just okay or happy about himself/herself when she is [action]?" (self-evaluation). For these two questions, the order of the response choices (e.g., all of the time vs. sometimes) was counterbalanced across trials.

Children were also asked, "How important do you think [action] is to [character]?" (self-relevance). To answer this question, children were shown a vertical 5-point Likert-type pictorial scale, with a series of yellow stars ranging in graduated size from ½ in. (bottom) to 2 in. in diameter (top). To ensure that each child understood the meaning of the scale points, familiarization with the star scale was done prior to the testing session. Specifically, children were told,

> "During my stories, we will use this star scale. You can use these stars to tell me how important you think things are to the kids in the stories. Let me show you how it works. Please tell me something you really like to do—something that means a whole lot to you. [child responds] Great! You really like X it means a lot to you—it is really important to you. So, let us point to the biggest star. [child points] Great! Now, tell me something you only like to do a tiny, little bit; something you do not care much about. [child responds]. OK! You only like X a tiny bit, it does not mean much to you—it is only a tiny, little bit important. So, let us point to the smallest star. [child points]. Great! Well, I like to go swimming sometimes—I do not care a whole lot about swimming, so it would not be the biggest star (point to the biggest star), but swimming is more than just a tiny, little bit important to me, so it would not be the smallest star (point to the littlest star). I think it is medium important to me, so I am going to point to the middle star (points to middle star)."

Also, each time the experimenter asked the self-relevance control question during testing, she reminded the child of the two scale anchors: "Remember, the biggest star means it is really, really important, and the littlest star means that it is only a tiny, little bit important."

For all story types, children were explicitly told the character's current desire (e.g., "Gloria really wants to paint pictures") and the prohibitive rule (e.g., "Gloria's mom says, 'You should not paint pictures'"). To verify that children followed this key information, the *rule* ("What rule does his/her mom say?") and *desire* control questions ("What does X want to do?") were asked in counterbalanced order across trials prior to the final scene.

Predict-and-explain test questions. At the conclusion of each story for the first six trials (two personal-essential, two personal-peripheral, two moral-essential), children were asked to predict and explain the character's *action decision*. They were provided with two forced-choice options: One action decision fulfilled the character's immediate desire (*transgression*) whereas the other action decision conformed to the parental rule (*compliance*). Picture cards illustrated each response choice. In these picture cards, characters were shown profile so that no emotion expression could be discerned from their face. As with Lagattuta (2005), willpower/compliance actions depicted the protagonist with his or her hands behind his or her back to show clearly that the character had decided not to engage in the prohibited activity.

As shown in Figure 1 (bottom right), for the Gloria personal domain restriction versions, children were asked, "Do you think Gloria will paint a picture or leave the paints alone?" As shown in Figure 2 (bottom row), the action decision test question for the moral rule version of the Gloria story asked, "Do you think Gloria will take her brother's paints or leave them alone?" The response choice options (e.g., paint a picture vs. leave the paints alone) were counterbalanced across story trials. Once the child chose an action decision, that card was placed in the story line to complete the sequence. The other response card was discarded. Next, the experimenter asked the child to explain the reason for the decision, for example, "So, Gloria chooses to X [child's response], why do you think Gloria chose to X? What was she thinking?" Because we were primarily interested in children's reasoning about what was going on inside the characters' minds, we only questioned participants about what the characters *would* do, not what they *should* do in that situation.

Final test questions for the predict-and-explain trials focused on the character's emotional response to the decision to fulfill his or her desire versus comply with a rule. Following Lagattuta (2005), children were asked a *primary emotion question*, for example, "How does Gloria feel right now, when she is [action]? Does she feel good or feel bad?" The order of the response choice was counterbalanced across story trials. The experimenter then asked the child to explain the reason for the emotion, as in, "Why does Gloria feel X?" Finally, children were asked the *secondary emotion probe*, "Could Gloria be feeling anything else right now, or just X? Why?"

Explain disobedience test questions. For the last two trials of the testing session, children were given two additional personal-essential stories. For these trials, instead of asking children to predict the character's action decision, the experimenter *told* participants that the character decided to fulfill his or her desire and then asked participants to explain why (the transgression decision card was added to the end of the story). Children were then asked to predict and explain the character's emotions following the same procedures as the predict-and-explain trials (i.e., the primary and secondary emotion probes). These explain disobedience trials focused on personal-essential stories because of our central interest in children's reasoning about situations where disobedience may indeed be legitimate. That is, the inclusion of these trials ensured that each child had, at minimum, two opportunities to explain acts of rule violation and to predict and explain emotions following noncompliance in situations where rules violated the personal domain and the desired action was essential to identity.

Three additional test probes were also included in the explain disobedience trials to gather additional information about how children reasoned about *rule violations* ("Did X follow his/her mom's rule or not follow it?"), *rule legitimacy* ("Do you think it was okay or not okay for [Gloria to paint a picture]?"), and *emotions* ("How do you think X would feel if he/she never, ever got to [paint] again?"). These probes were not included in the initial six predict-and-explain trials because we did not want to influence children's responses by calling attention to issues of the legitimacy or acceptability of different kinds of rules.

General Procedure

Participants were interviewed individually in a quiet room by a female experimenter after informed consent was provided by the university research ethics board, school authorities, and the child's parents. Before the interview began, the experimenter also asked children to provide verbal assent. Each participant received a total of eight story trials: six predict-and-explain trials (two personal-essential, two personal-peripheral, and two moral-essential) followed by two explain disobedience trials (personal-essential). To maximize children's comprehension of the peripheral versus essential story premise, the experimenter repeated the relevant story element(s) if children made any errors in their responses to the control questions. For example, if the child stated that a character in a peripheral version of a story liked to engage in the activity ''all of the time'' they were reminded both verbally and via pointing to the story pictures that the character did it ''only sometimes.'' The procedure lasted approximately 30 min.

Due to the repeated measures design, several measures were taken to control for potential order effects. Personal-peripheral trials always appeared as the first two trials because these were the only trials where the character had a peripheral interest in the desired activity (all other trials had the essential core). Initial pilot testing revealed that children attended less to the unique premise of the peripheral versions when they appeared after the essential story versions. The presentation of the moral-essential and personal-essential predict-and-explain trials (trial order 3–6) was counterbalanced, with each story type appearing equally in each trial order slot, and with trials of the same story type (e.g., moral-essential) never appearing consecutively. No participant received more than one trial of a story type with the same desire focus (e.g., receiving both the ''sparkly clothes'' and ''superhero clothes'' stories for the two personal-peripheral trials). Finally, all specific stories + ending types (e.g., Gloria moral-essential) appeared equally for each age group. Twenty unique order combinations were created, with one version randomly assigned to each participant in each age group. As stated previously, all forced-choice response options for control questions and test questions were counterbalanced across trials so that children with a response bias to pick the first or second option would respond at chance.

Coding of Explanations for Decisions and Emotions

Trained undergraduate research assistants independently coded the behavior and emotion explanations from verbatim written transcripts of the experimental sessions. Coders were blind to participant-identifying information. Behavior and emotion explanations were categorized into three main categories following criteria by Lagattuta (2005): *goal oriented*, *rule oriented*, and *future oriented*. Cohen's Kappa based on 75% of the transcripts was high, $\kappa = .96$.

Goal-oriented explanations fit within a theory of mind, desire psychology framework. They explained the behavior or emotion as caused by the character's personal goals or desires (e.g., ''Because she wanted to do what she wanted to do,'' ''Because he wanted to wear his superhero clothes''). Rule-oriented explanations followed more of a sociomoral perspective. They explained the behavior or emotion as caused by adherence, failure, or obligation to follow the rules (e.g., ''Because he listened to his mom,'' ''Because she broke the rule''). Finally, future-oriented explanations explained the emotion or behavior as caused by a potential future outcome or consequence (e.g., ''Because his mom will get mad,'' ''Because she could have hurt that girl if she pushed her''). Future-oriented explanations cross-cut theory of mind, social domain theory, and self perspectives in that they provide insight into children's reasoning about how a person's decisions and emotions in the here-and-now are influenced by his or her expectations for how those actions will impact the self or another person in the future (see Atance & O'Neill, 2005; Lagattuta, 2005, 2007, 2008). Future-oriented explanations were further coded as to whether they described consequences to the self, another person, or both the self and another person.

There were two additional coding dimensions unique to this study: *identity-related* and *moral-evaluative* explanations (Cohen's $\kappa = .95$). As described, these categories were of theoretical interest because we wanted to examine whether children would: (a) more often refer to aspects of self or identity when explaining how characters would respond to personal domain restrictions versus moral rules and (b) more frequently explain behavioral or emotional responses to moral rules versus personal domain restrictions in relation to how others or the self would evaluate the behavior (e.g., as right or wrong, good or bad).

Identity-related explanations explained the behavior or emotion as caused by the importance of the activity to the character's identity (''Because that's part of who he is,'' ''Because she always does it''), including references to stable attitudes or preferences (e.g., ''Because he loves to play with that friend,'' ''Because she doesn't care about painting''). To be conservative, identity-related explanations excluded simple mentions of the character's immediate desires. Desires, even though self-relevant, do not clearly indicate enduring attitudes. That is, whereas attitudes such as *like, love,* and *hate* (see Heider, 1946) convey dispositional stability (a lasting positive or negative orientation toward something), "wants" are often transitory or situational, and they provide no clear indication that the preference did or will persist over time (Gelman, 2004). Therefore, the explanation "Because she wanted to paint" would not be considered an identity-related explanation; however, "Because she likes to paint" would be coded as such. Moral-evaluative explanations explained the behavior or emotion as caused by an actual or anticipated emotional reaction or moral evaluation by either another person or by the self (e.g., "'Because he thinks it was a bad thing to do," "Because his mom will be really disappointed," "Because stealing is wrong").

As with Lagattuta (2005) explanations could be coded into one or multiple categories. The coding categories were not mutually exclusive. For example, the explanation, "[Gloria feels bad] because her mom said that she wasn't allowed to paint and she really, really wanted to paint a picture," was coded as both goal oriented and rule oriented.

Results

Results are organized into four sections. Preliminary analyses examine children's answers to the control questions regarding the story premises. We then analyze children's responses to the predict-and-explain trials: first their predictions and then their explanations for story character's decisions and emotions. Finally, we analyze children's responses for the explain disobedience trials. Except when otherwise noted, data were analyzed using repeated measures analysis of variance (ANOVA), with the significance level set at $p < .05$. Significant main effects were followed up with Tukey's honestly significant difference (HSD) tests to evaluate pairwise comparisons, with only significant pairwise comparisons among means reported. Simple effects tests examined significant interactions. Partial eta squares (η_p^2) are provided as a measure of effect size for all significant main effects and interactions.

Control Questions

Repeated measures ANOVAs assessed whether children distinguished the essential core stories from the peripheral core stories. Three story types (personal-essential explain disobedience, personal-essential, moral-essential) had essential cores, whereas one story type (personal-peripheral) had the peripheral core. Scores ranged from 0 to 2 per story type totaling the number of trials children answered each forced-choice question correctly the first time. Correct answers for the essential core stories included: Character liked to do the activity "all of the time" and felt "happy about self" when participating in that activity. Correct answers for the peripheral core stories included: Character liked to do activity "some of the time" and felt "just OK about self" when engaged in that action. For the "how important" question, we averaged ratings for the star scale (from 1 = *tiny, little bit important* to 5 = *very important*) across the two trials of each story type. Preliminary analyses showed that the order of the response options did not affect answer choices. All age groups utilized the 5 points of the star scale, and all participants correctly answered the desire and rule control questions.

In response to whether the character liked to do the activity "all of the time" versus "some of the time," a 3 (age: 4, 5, 7) × 4 (story type: personal-essential explain disobedience, personal-essential, moral-essential, personal-peripheral) repeated measures ANOVA resulted only in a main effect for age, $F(2, 52) = 3.25$, $p < .05$, $\eta_p^2 = .11$. Four-year-olds ($M = 1.69$) gave correct responses significantly less often than 7-year-olds ($M = 1.91$), $p < .05$, Tukey's HSD (M for 5-year-olds = 1.80). Age and story type main effects emerged for correct responses to whether the character "felt happy versus just OK" when doing the activity: $F(2, 54) = 4.95$, $p < .01$, $\eta_p^2 = .16$ for age, and $F(3, 162) = 6.86$, $p < .0001$, $\eta_p^2 = .11$ for story type. Participants more often correctly responded to the three essential stories ($Ms = 1.85$, 1.68, and 1.67, respectively) than to the peripheral stories ($M = 1.57$). Four-year-olds made more errors ($M = 1.50$ correct) than 7-year-olds ($M = 1.90$), $p < .01$ (M for 5-year-olds = 1.68). Still, participants rated the activities in the essential stories ($Ms = 4.80$, 4.79, and 4.79, respectively) as significantly higher in importance to the character than the activities in the peripheral stories ($M = 2.95$), $F(3, 162) = 160.04$, $p < .0001$, $\eta_p^2 = .75$. Overall, then, children responded appropriately to most control questions, with 4-year-olds showing some difficulty with the premise of the personal-peripheral stories. As described, the experimenter corrected control question errors before continuing by repeating back the relevant story elements and providing the correct response.

Predictions for Decisions and Emotions: Predict-and-Explain Trials

For the predict-and-explain trials, four predictions are of central interest: (a) the frequency that children predicted the character would abide by the rule (*predict compliance*), (b) the frequency that children predicted that the character would comply with the rule and feel good as the primary emotion (*feel good compliance primary*), (c) the frequency that children predicted that the character would abide by the rule and feel good as the primary or secondary emotion (*feel good compliance total*), and (d) the frequency that children predicted that the character would comply and feel good (feel good compliance) or transgress and feel bad (*feel bad transgression*) as the primary or secondary emotion: a broader measure of moral-normative responses. Scores ranged from 0 to 2 totaling the number of trials per story type that children provided each type of response. Preliminary analyses showed that children's predictions were not influenced by (a) gender of story character, (b) the order of the desire versus rule control questions, (c) the order of the action decision choices, or (d) the type of activity (clothing, friend, sport activity, or art activity choice). These factors were removed in the analyses that follow.

Predict compliance. As shown in Figure 3, a 3 (age) × 2 (sex) × 3 (story type: personal-essential, personal-peripheral, moral-essential) repeated measures ANOVA for predict compliance scores resulted in a main effect for story type, $F(2, 108) = 3.62$, $p < .03$, $\eta_p^2 = .12$, qualified by a significant Age × Story Type interaction, $F(4, 108) = 3.81$, $p < .03$, $\eta_p^2 = .12$. Children judged that story characters would comply with rules significantly more often for moral-essential trials ($M = 1.07$, $SE = .11$) than for personal-essential trials ($M = .80$, $SE = .11$), $ps < .05$. Simple effects tests, however, revealed that story type was only significant for the 7-year-olds. In addition, age was a significant factor only for moral-essential stories, with 7-year-olds predicting compliance with moral rules ($M = 1.50$, $SE = .19$) significantly more often than 4- ($M = .95$, $SE = .19$) or 5-year-olds ($M = .75$, $SE = .19$), $ps < .05$.

Predict feel good compliance as primary emotion. A 3 (age) × 2 (sex) × 3 (story type) repeated measures ANOVA for predict feel good compliance primary scores resulted in no significant main effects or interactions. All age groups rarely predicted that story characters would comply with the rule and feel good as the primary emotion (4-year-olds: $M = .23$, $SE = .10$; 5-year-olds: $M = .28$, $SE = .10$; 7-year-olds: $M = .40$, $SE = .10$).

Predict feel good compliance as primary or secondary emotion. Figure 4 shows data for children's decision-emotion predictions in response to the primary or secondary emotion probe. Of interest was whether children ever predicted that the person would comply and feel positive emotions. A 3 (age) × 2 (sex) × 3 (story type) ANOVA resulted in a main effect for story type, $F(2, 108) = 6.14$, $p < .01$, $\eta_p^2 = .19$, qualified by an Age × Story Type interaction, $F(4, 108) = 3.42$, $p < .04$, $\eta_p^2 = .11$. Simple effects tests showed that 7-year-olds predicted feel good compliance significantly less often for personal-essential stories ($M = .30$, $SE = .13$) than for personal-peripheral ($M = .60$, $SE = .15$) or moral-essential ($M = .85$, $SE = .17$) stories, $ps < .05$. Younger children did not differentiate by story type. Age differences only emerged in predictions for the moral-essential situations, with 7-year-olds predicting feel good compliance ($M = .85$, $SE = .17$) significantly more often than 4- ($M = .30$, $SE = .17$) or 5-year-olds ($M = .45$, $SE = .17$), $ps < .05$.

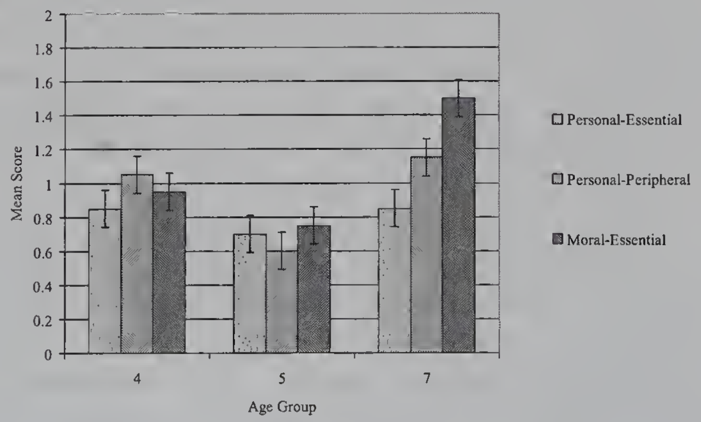


Figure 3. Mean scores (+*SE*) for predicting compliance.

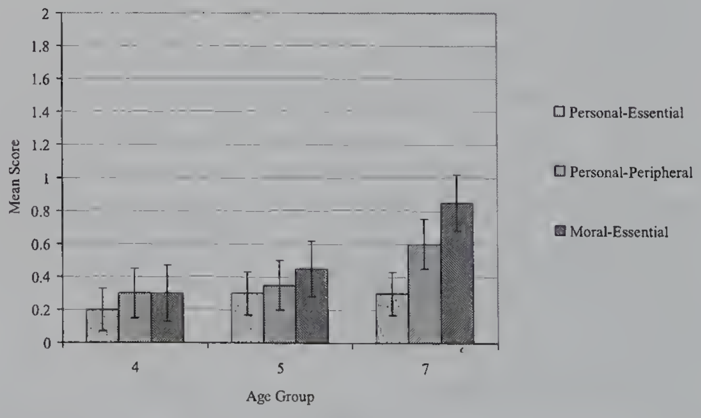


Figure 4. Mean scores (+*SE*) for predicting compliance + feel good as the primary or secondary emotion.

Predict feel good compliance or feel bad transgression. Final prediction analyses examined children's more global provision of moral-normative responses: predicting feel good compliance or feel bad transgression as either the primary or secondary emotion. As shown in Figure 5, a 3 (age) × 2 (sex) × 3 (story type) repeated measures ANOVA resulted in significant main effects for age, $F(2, 54) = 3.91$, $p < .02$, $\eta_p^2 = .13$, and for story type, $F(4, 108) = 5.32$, $p < .01$, $\eta_p^2 = .17$. Five- ($M = 1.08$, $SE = .14$) and 7-year-olds ($M = .95$, $SE = .14$) made these moral-normative predictions significantly more often than 4-year-olds ($M = .57$, $SE = .14$), $p < .05$. Moral-essential stories provoked a significantly higher frequency of feel good compliance or feel bad transgression predictions ($M = 1.08$, $SE = .11$) compared to either the personal-essential stories ($M = .70$, $SE = .10$) or personal-peripheral stories ($M = .82$, $SE = .11$), $ps < .05$.

Interim summary. Seven-year-olds exhibited greater attention to rule type and identity relevance than 4- and 5-year-olds when making predictions about people's decisions and emotions. Analyses examining children's predictions of "compliance" as well as "feel good compliance" (total) showed that only 7-year-olds differentiated by story type: They more frequently provided these responses to moral-essential versus personal-essential trials. In contrast, 4- and 5-year-olds predicted low rates of compliance and feel good compliance across trials. Additional prediction analyses encompassing a broader range of moral-normative responses (predictions of feel good compliance *or* feel bad transgression) revealed sensitivity to rule type and identity relevance even in younger children. That is, all age groups most often judged that characters would abide by rules and feel positive emotions or transgress and experience negative emotions in situations involving moral rules, and they least often made such predictions in situations involving personal domain restrictions that forbade activities essential to the character's identity.

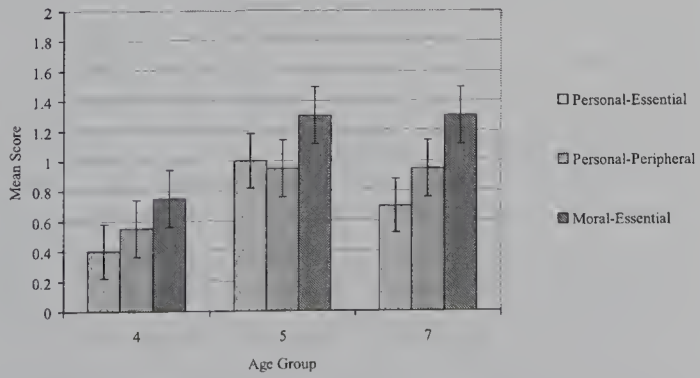


Figure 5. Mean scores (+*SE*) for predicting feel good compliance or feel bad transgression.

Explaining Action Decisions and Emotions: Predict-and-Explain Trials

Explanation analyses examine children's provision of goal-, rule-, and future-oriented explanations for story characters' decisions and emotions, as well as the frequency that children explained behaviors or emotions by making references to the character's identity or to moral-evaluative concerns (see the Method section for coding criteria). Scores ranged from 0 to 2, totaling the number of trials per *each story type* that children provided that type of explanation to justify *either* the character's action decision *or* emotion (regardless of what they predicted the character would do or feel). This scoring procedure reduces reporting redundancy because preliminary analyses on the behavior and emotion explanations separately yielded the same results. Because children did not predict the character's action decision in the explain disobedience trials (they were told the character disobeyed), these data are presented separately in a subsequent section.

Goal-oriented explanations As shown at the top of Table 2, a 3 (age) × 2 (sex) × 3 (story type) repeated measures ANOVA for goal-oriented explanations resulted in no main effects or interactions. Children frequently explained behavior choices and emotions in relation to character's desires, intentions, or goals regardless of story type ($Ms = 1.38$ for 4-year-olds, 1.62 for 5-year-olds, and 1.65 for 7-year-olds).

Rule-oriented explanations As shown in the middle of Table 2, a 3 (age) × 2 (sex) × 3 (story type) repeated measures ANOVA for rule-oriented explanations resulted in a main effect for age, $F(2, 54) = 5.57$, $p < .01$, $\eta_p^2 = .17$, and a near-significant main effect for sex, $F(1, 54) = 3.56$, $p = .065$, $\eta_p^2 = .06$. Seven-year-olds provided rule-oriented explanations for characters' decisions and emotions ($M = 1.62$, $SE = .14$) significantly more often than 4- ($M = .98$, $SE = .14$) and 5-year-olds ($M = 1.10$, $SE = .14$), $ps < .05$ (Tukey's HSD). There was a trend for females to explain actions and feelings in relation to rules more often than males ($Ms = 1.39$ vs. 1.08).

Future-oriented explanations The bottom of Table 2 displays data for future-oriented explanations. A 3 (age) × 2 (sex) × 3 (story type) repeated measures ANOVA revealed significant main effects for story type, $F(2, 108) = 7.06$, $p < .001$, $\eta_p^2 = .12$, and for age, $F(2, 54) = 4.68$, $p < .01$, $\eta_p^2 = .15$. Children explained behaviors and emotions in relation to potential future consequences significantly more

Table 2

Emotion Explanations Scores by Age and Story Type for Predict-and-Explain Story Trials[a]

	Age group		
Emotion explanation type	4	5	7
Goal-oriented explanations (total)	1.38	1.62	1.65
Personal-essential	1.50 (.69)	1.70 (.66)	1.70 (.57)
Personal-peripheral	1.30 (.80)	1.55 (.76)	1.65 (.59)
Moral-essential	1.35 (.75)	1.60 (.68)	1.60 (.75)
Rule-oriented explanations (total)	0.98	1.10	1.62
Personal-essential	0.95 (.95)	1.10 (.85)	1.55 (.76)
Personal-peripheral	1.10 (.79)	0.95 (.89)	1.75 (.64)
Moral-essential	0.90 (.85)	1.25 (.79)	1.55 (.76)
Future-oriented explanations (total)	0.08	0.30	0.53
Personal-essential	0.00 (0)	0.25 (.64)	0.30 (.57)
Personal-peripheral	0.05 (.22)	0.30 (.66)	0.45 (.69)
Moral-essential	0.20 (.52)	0.35 (.67)	0.85 (.75)

[a]Mean scores out of two trials per participant. Data represent the number of trials each type of explanation was provided for either the character's behavior decision and/or emotion. Standard deviations are shown in parentheses.

often for moral-essential stories ($M = .47$, $SE = .09$) than for all other story types, and significantly less often for personal-essential stories ($M = .18$, $SE = .07$) than for all other story types, $ps < .05$. Seven-year-olds explained characters' reactions in relation to potential future outcomes ($M = .53$, $SE = .10$) significantly more often than 4-year-olds ($M = .08$, $SE = .10$), $p < .05$ (Tukey's HSD). Descriptively, 44% of future-oriented explanations focused just on outcomes for the self, 16% focused just on potential consequences for another person, with the remaining 40% referring to possible outcomes for both self and other.

Identity-related and moral-evaluative explanations Final analyses on the predict-and-explain trials focused on the number of trials per story type that children explained characters' action decisions and resulting emotions in relation to the characters' sense of identity (e.g., what he or she likes to do, what is important to the character) or by citing moral-evaluative concerns (e.g., what other people might think of him or her; whether it was a good or bad thing to do). As shown in Figure 6, a 3 (age) × 2 (sex) × 3 (story type) repeated measures ANOVA for identity-related explanations resulted in a significant main effect for story type, $F(2, 108) = 9.65$, $p < .0001$, $\eta_p^2 = .15$. All age groups used identity-related explanations significantly more often to justify characters' decisions and emotions in situations involving personal domain restrictions—both personal-essential ($M = .92$, $SE = .11$) and personal-peripheral ($M = .72$, $SE = .12$)—as opposed to moral rules ($M = .43$, $SE = .09$), $ps < .05$.

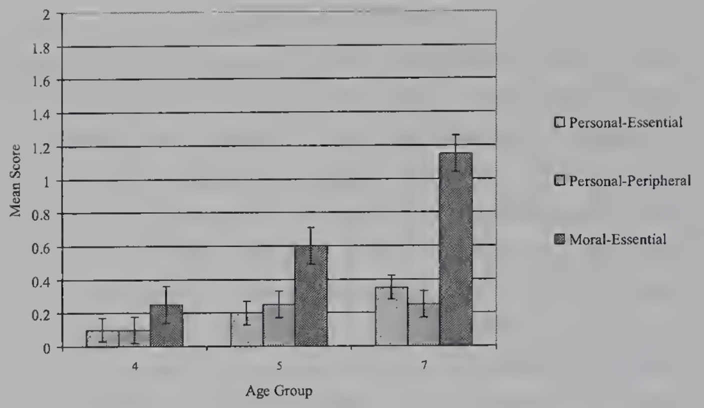


Figure 6. Identity-related explanations (+*SE*) by story type.

The opposite pattern arose for moral-evaluative explanations. As shown in Figure 7, a 3 (age) × 2 (sex) × 3 (story type) repeated measures ANOVA for moral-evaluative explanations resulted in main effects for story type, $F(2, 108) = 20.40$, $p < .001$, $\eta_p^2 = .27$, and for age, $F(2, 54) = 5.99$, $p < .004$, $\eta_p^2 = .18$; both qualified by a Age × Story Type interaction, $F(4, 108) = 4.27$, $p < .003$, $\eta_p^2 = .14$. All age groups explained characters' reactions in relation to moral-evaluative concerns significantly more often for the moral-essential stories ($M = .67$, $SE = .08$) than for either type of personal domain restriction story ($Ms < .22$, $SEs < .07$), $ps < .05$. Age differences only emerged for the moral-essential stories, with 7-year-olds providing moral-evaluative explanations ($M = .58$, $SE = .09$) significantly more often than 5-year-olds ($M = .35$, $SE = .09$), and 5-year-olds doing so more often than 4-year-olds ($Ms = .15$, $SE = .09$), $ps < .05$.

Explain Disobedience Trials

Analyses for the two explain disobedience trials focus on the following responses: (a) predicting positive emotions for transgression as either the primary or secondary emotion, (b) provision of identity-related and moral-evaluative explanations, (c) stating that the character did not obey the rule, (d) judging that disobeying the rule was "not OK," and (e) predicting that the character would feel bad if he or she could never engage in that action. Scores ranged from 0 to 2 summing the number of trials per story type that children provided each type of response.

First, consider children's predictions that the character would "feel bad" after disobeying a rule

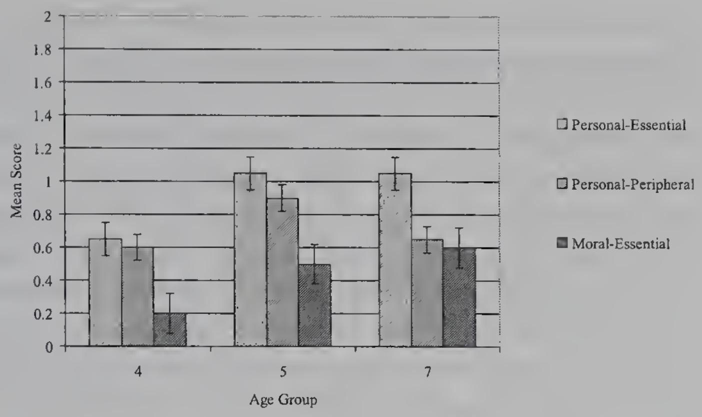


Figure 7. Moral-evaluative explanations (+*SE*) by story type.

that forbade participation in a personal domain activity essential to the character's sense of identity. A 3 (age) × 2 (sex) ANOVA for primary emotion predictions resulted in no main effects or interactions. A parallel ANOVA for predicting negative emotions as either the primary or secondary emotion also resulted in no main effects or interactions. All age groups infrequently predicted that transgressing against rules in the personal domain would lead to negative emotions as the primary emotion (*M*s = .15, .45, .30 for 4-, 5-, and 7-year-olds, respectively) or as either the primary or secondary emotion (*M*s = .25, .65, and .70).

Children's explanation patterns for the personal-essential explain disobedience trials paralleled the findings from the personal-essential predict-and-explain trials: Identity-related explanations were provided frequently by participants as a justification for the characters' decisions or emotions (*M*s = 1.20 trials for 4-year-olds, 1.35 for 5-year-olds, and 1.30 for 7-year-olds), whereas moral-evaluative explanations were nearly nonexistent (*M*s = .35, .25, and .20). Univariate ANOVAs confirmed there were no age differences, *p*s > .05.

We then examined children's beliefs that a rule had, in fact, been broken. A 3 (age) × 2 (sex) ANOVA for disobedience scores resulted in a main effect for age, $F(2, 54) = 4.38$, $p < .017$, $\eta_p^2 = .14$. Whereas 5- and 7-year-olds judged that a rule had been broken 100% of the time, 4-year-olds did so only 85% of the time. Moreover, most children judged that breaking the rule was "not OK" to do (*M*s = 1.65, 1.60, and 1.75 for 4-, 5-, and 7-year-olds, respectively). A 3 (age) × 2 (sex) ANOVA for "not OK" scores resulted in no main effects or interactions.

Finally, children predicted nearly 100% of the time (*M*s = 1.95 for 4-year-olds, 2.0 for 5- and 7-year-olds) that the characters would feel bad if they never could participate again in the desired activity that was essential to their identity.

Discussion

At the same time that children are learning to comply with societal and parental rules, they begin to carve out a personal domain of actions where they can actively stake claims to autonomy, control, and choice (Nucci & Turiel, 2000). Put simply, children learn to identify situations where there may be legitimate grounds for *disobeying*. Although previous research has established that older children and adolescents acknowledge that these claims to the personal domain are important for autonomy and individual identity (Nucci & Turiel, 2000), little is known about the emergence of these intuitions during early childhood. The current study provides a new approach by bridging methodology from social domain theory (Smetana, 2006), essentialism (Gelman, 2003), and theory of mind (Lagattuta, 2005). Findings reveal that when young children reason about people's decisions and emotions, they attend to rule domain (personal vs. moral) and identity relevance (essential vs. peripheral). Data are most robust for 7-year-olds, with evidence for initial insights in children as young as 4 years.

When Will a Person Decide to Comply?

When making predictions about what people will do in situations where their desires conflict with rules, does desire trump rule or rule trump desire? As predicted, there were no significant age differences between 4 and 7 years in predicting that people would comply with rules that intruded on the personal domain: Participants predicted that characters would disobey parents and fulfill their desires the majority of the time (e.g., Carlos will play baseball, Tom will play with his best friend, Sarah will wear her sparkly clothes). In contrast, when situations involved moral rules, predictions of compliance significantly increased with age (nearly double the rate), even though the desired activity was described as essential to that person's identity (e.g., "Max the LegoBuilder" will not steal to build with highly desired Legos). Sensitivity to rule type and identity relevance was clearest for the 7-year-olds: They judged that characters would comply with rules significantly more often for moral-essential than for personal-essential trials (personal-peripheral trials did not differ from either type). In contrast, 4- and 5-year-olds did not differentiate by story type and typically predicted that desires would drive people's behaviors. Future research using open-ended (What do you think Gloria will do?) versus forced-choice (Do you think Gloria will X or Y?) prompts would also be revealing because it would allow children the opportunity to suggest compromise solutions (e.g., comply now and paint later). Analyses of such compromise suggestions could provide further insight into how young children learn to establish the boundaries of a personal domain via negotiating and making concessions with authority figures.

Previous research has shown that children as young as 3 years view moral rules as more obligatory than rules governing the personal domain and treat moral rule infractions as more severe than breaking rules that intrude on the personal domain (Killen & Smetana, 1999; Nucci, 1981; Tisak, 1993). Thus, it is puzzling why 4- and 5-year-olds did not predict higher rates of compliance with moral rules versus personal domain restrictions in the current study. Still, past researchers have not asked children to judge whether people will act to fulfill desires essential to their identity versus abide by moral rules. Rather, studies have focused on children's reasoning about whether children versus parents (or teachers) have the right to make certain rules, whether certain rules can be changed, how severely a person should be punished for breaking certain rules, or whether a parent versus child should make decisions (e.g., Nucci & Weber, 1995). These are different questions than whether they believe highly intense desires versus rules more strongly motivate behavior. Given that children develop early knowledge (age < 2 years) that desires drive actions (see Wellman & Lagattuta, 2000), they likely have difficulty factoring other aspects of the situation—in this case, the importance of rules or possible harm to others—that may alter this relation (see Lagattuta, 2005, 2008). Potentially, moral-peripheral situations, a story type not tested in the current study, may be a situation where 4- and 5-year-olds may predict higher compliance because in these cases the desired activity is less important.

Does Abiding by Rules Feel Good and Transgressing Feel Bad?

Research on young children's reasoning about emotions and rules has focused on children's understanding that harming others for personal gain can make the victimizer feel bad. Results have consistently found that whereas children 7–8 years and older acknowledge that victimizers can experience negative emotions, younger children typically judge that victimizers just feel happy (see Arsenio et al., 2006). Recently, Lagattuta (2005) identified the same age shift for children's judgments about social-conventional and prudential rules: With age, children increasingly predict that compliance can feel good (despite desire inhibition) and transgression can feel bad (despite desire fulfillment). The current research takes a different approach. Importantly, predictions of feel good compliance and feel bad transgression should not always be considered the optimal, or most developmentally advanced response. In some situations, notably when authority figures restrict actions essential to sense of self or identity, judgments of "feel bad compliance" and "feel good transgression" may be more appropriate.

Although all age groups understood that characters would feel bad being forbidden from activities essential to self, 7-year-olds showed the strongest ability to differentiate situations by rule domain and identity relevance. Seven-year-olds predicted feel good compliance at significantly lower rates for the personal-essential scenarios than for all other story types, and they predicted significantly higher rates of feel good compliance for the moral-essential trials than all other story types. Age differences only emerged for the moral-essential scenarios: Seven-year-olds judged that compliance with moral rules can make a person feel good significantly more often than did 4- and 5-year-olds. Story type did not affect 4- and 5-year-olds' predictions of feel good compliance: Young children infrequently predicted these action–emotion sequences. Importantly, however, when analyses broadened to include both predictions of feel good compliance and feel bad transgression, all age groups treated moral rules differently than personal domain restrictions. Thus, although there is some early awareness about different action–emotion sequences depending on rule domain and identity relevance, between the ages of 4 and 7 children demonstrate increasing ability to differentiate by situation when reasoning about people's decisions and emotions.

These developmental findings correspond with data by Shaw and Wainryb (2006) showing that 7-year-olds (as with older, but not younger children) judge that not all compliance is good: Going along with unfair peer demands can cause negative emotions, lower self-worth, and increased risk of further victimization. More broadly, children's responses resonate with Werner and Kaplan's (1963) theory of increased differentiation and integration of concepts over development. That is, 4- and 5-year-olds (compared to 7-year-olds) had more difficulty differentiating and integrating information about rule domain and identity relevance to reason about how the character would act and feel (see also Helwig, Zelazo, & Wilson, 2001).

Clues from Children's Explanations

Children's explanations for why characters made certain decisions and experienced particular emotions clarify and extend the prediction data. Foremost, they reveal that children as young as 4 years do attend to rule domain and identity relevance when reasoning about people's actions and emotions. Unlike most analyses on the prediction data, significant main effects for story type in the explanation data were not qualified by significant interactions with age. Across age, children provided future-oriented (e.g., "She maybe felt bad for the girl she hit because maybe she was crying") and moral-evaluative (e.g., "It wouldn't be very nice to take somebody else's clothes and just leave, because they're the other person's") explanations significantly more often for moral-essential scenarios than for situations featuring personal domain restrictions. These data support previous research showing that young children make distinctions between the moral and personal domains in relation to whether the action could have harmful consequences for others (Killen & Smetana, 1999; Nucci, 1981). Developmental patterns also converge with prior studies. Seven-year-olds provided future-oriented and moral-evaluative explanations significantly more often than 4- and 5-year-olds, fitting with previous research showing increased attention to and understanding of future possibilities (Atance & O'Neill, 2005; Lagattuta, 2005, 2007), as well as rising concern about being valued, accepted, and approved by others during this age period (Higgins, 1991; Ruble & Frey, 1991).

In contrast, personal-essential situations elicited significantly more justifications focusing on aspects of the character's sense of self and identity (e.g., "Because she loves playing soccer so much . . . that's part of who she is") compared to personal-peripheral and moral-essential story types. Thus, children showed awareness of connections between identity and the need to stake claims to a personal domain of action and choice. Here is one particularly revealing example from a 7-year-old: "'Cause she thinks that Sue is such a nice person that she wouldn't, she wouldn't be herself if she didn't get to play with her anymore, cause she likes to play with her so much that if she didn't do it, then she would be so sad all the time, because she wants to play with Sue so bad." Here again, it is important to stress that for both the personal-essential and the moral-essential scenarios, the desire to participate in the action was very high. Thus, although children provided identity-related explanations based on the intensity of the desire (personal-peripheral < personal-essential), they also attended to rule domain (moral-essential < personal-peripheral < personal-essential).

Two types of explanations did not vary by story type: goal oriented and rule oriented. This is understandable given that these explanations could sensibly be applied to compliance (e.g., "Because she didn't get what she wanted" [goal oriented], "Because she followed the rule" [rule oriented]) or transgression decisions (e.g., "Because he wanted to wear those clothes and he did it," "Because she didn't listen to what her mom said"). Nevertheless, as with previous research (Lagattuta, 2005), rule-oriented explanations did significantly increase with age, showing increased ability to integrate theory of mind and social-normative perspectives: People's emotions and decisions are not just motivated by desires and goals but also by attention to rules.

Issues and Unanswered Questions

Did children understand that a rule had been broken? As even 7-year-olds most often predicted that characters would disobey personal domain restrictions, perhaps children did not treat these restrictions as "rules." Methods and data argue against this interpretation. All story trials explicitly stated the rule (X's Mom says, "You should not . . .") and included a rule control question ("What *rule* did his/her mom say?"). For the explain disobedience trials (featuring personal-essential situations), children were directly asked whether the story character broke the mother's rule and whether it was OK to do so. All age groups correctly answered the rule control question for 100% trials. Contrary to our predictions, there was not an age-related decline in judgments that disobeying personal domain restrictions was "not OK": All age groups, even 7-year-olds, stated that these acts of noncompliance were "not OK" (*Ms* = 83%–88% trials).

This intriguing disconnect between children's judgments—yes, the character will break the rule and feel good, but no, it is not OK to make this decision—has been found in previous studies. For example, Lourenço (1997) reported that although the majority of 4- and 6-year-olds predicted a person who broke a moral rule would feel happy, they acknowledged that the victimizer was "bad." Moreover, Kalish and Shiverick (2004) found that whereas 30% of children said that the character would actually obey a rule, 75%–95% said he or she should follow the rule (see also Nucci & Weber, 1995). Thus, children as young as 4 years understand that people knowingly—not naively—decide to disobey rules. Given that children make even stronger demands for personal jurisdiction during adolescence (Nucci, 2006), future research may find that this disconnect changes over time to *convergence*: Yes, he will break the rule, yes, he will feel happy, and yes, it is OK to make that decision (e.g., because the mother did not have the right to make that rule in the first place).

Did young children understand the essential versus peripheral distinction? In our procedures, 7-year-olds demonstrated greater knowledge than 4- and 5-year-olds about the impact of identity relevance on people's decisions and emotions. One interpretation is that because older children more highly attend to people's characteristics, attributes, and personality traits (Harter, 1998; Ruble & Frey, 1991), they have more advanced knowledge about the role identity plays in people's decisions and emotions. Another possibility, however, is that developmental differences emerged because 4- and 5-year-olds failed to understand the initial premise of the peripheral stories—that the activity, although desired at this moment, was not a central part of that character's sense of self. Although all age groups rated the activities higher in importance to characters when actions were described as essential versus peripheral to their identity, young children made more mistakes than 7-year-olds in answering the control questions for the peripheral core stories. Thus, because young children did not demonstrate equal initial understanding of the peripheral versus essential premise, it may have affected their ability to differentiate by identity relevance when reasoning about people's actions and emotions.

A challenge for future research, then, is to make the essential versus peripheral distinction more salient to young children. For the current study, we purposely featured actions within the personal domain (e.g., playing soccer, building with Legos) that are popular with children of this age group. As a result, children may have had difficulty imagining that another person would not enjoy these activities. One potential solution is to use novel-label predicates instead of known actions (e.g., "X likes to flimmer" and the rule is "not to flimmer"; see Kalish & Shiverick, 2004). A substantial drawback to this approach, however, is that when a novel label is used, the child has no information from which to judge whether the action is in the personal, moral, or social-conventional domain. For example, the child cannot ascertain whether by choosing "to flimmer" he or she could harm others. The current data reveal that this distinction—between moral rules versus personal domain restrictions—has an even higher impact on children's reasoning about people's decisions and emotions than identity relevance. Thus, if future studies use novel labels for actions, it would be critical to provide clues to rule domain: "X wants to flimmer. If X flimmers, someone else could get hurt" (moral rule) or, "X wants to flimmer. If X flimmers, X could get hurt" (prudential rule).

Do these findings apply more broadly to other social classes or cultures? Results of the current study are based on the reasoning of middle-class American children. Because these children typically have relative freedom to choose what to wear, who to play with, and what to do (see Kusserow, 2004), they may develop earlier knowledge about the emotional costs of complying with personal domain restrictions. Research has shown that children and adolescents in other cultures have different ideas about rights to autonomy and democracy in decision-making processes (see Helwig, 2006; Helwig, Arnold, Tan, & Boyd, 2007), and this could shape their reasoning about decisions and emotions in rule situations. Still, although the particulars of what gets defined as the personal domain can vary across cultural settings (e.g., Muslim women cannot choose what to wear), the establishment of a zone of personal choice and privacy appears to be culturally universal. Claims to personal jurisdiction have been documented in diverse cultures and social classes in North America, South America, the Middle East, and Asia (see Nucci, 2001; Smetana, 2002, 2006). These cross-cultural findings strongly suggest that our core data would likely generalize beyond the sample tested given cultural sensitivity to which particular activities are defined as personal.

Adult–Child Interactions and Socialization

Young children's judgment that people frequently disobey parental rules that intrude on the personal domain—and feel good about breaking these rules—has implications for adult–child interactions and socialization practices. Foremost, they argue for balance in promoting morality in young children: not only restricting actions that they should not do but also helping them to identify situations where they *can* assert personal control (see Nucci, 2006). Such an approach does not advocate telling children to blatantly disobey authority but rather promotes helping children respectfully negotiate socially and culturally acceptable areas of personal choice. Negotiation of claims to a personal domain has been found across very diverse cultures and social classes, with adults granting some space, and children seeking control to safeguard their developing identity and autonomy (Grusec & Goodnow, 1994; Nucci, 1994; Nucci, Camino, & Sapiro, 1996; Smetana, 2002). The current findings show that by 4 years of age children have some appreciation about these connections between identity development, and personal control. Thus, overregulation of the child's personal domain may be psychologically harmful in that the adult not only restricts the child the ability to express his or her self but also evaluates that aspect of the child's identity as immoral or unworthy (e.g., "You shouldn't act that . . . dress like that . . . be friends with them!"). Recent evidence supports these mental health implications. Adolescents whose parents severely restrict control over a personal domain report elevated rates of anxiety and depression, a finding documented in the United States and Japan (Hasebe, Nucci, & Nucci, 2004) as well as in urban and rural China (Helwig, Yang, Nucci, Yun, & To, 2009).

Conclusions

Results of the current research show that between the ages of 4 and 7 years children develop emerging knowledge about acts of legitimate disobedience: There are certain situations where people will resist authority and feel good about making that decision. All age groups infrequently predicted feel good compliance for situations featuring personal domain restrictions, with 7-year-olds least often doing so when rules forbade actions essential versus peripheral to the characters' identities. Moreover, all age groups more often justified characters' actions and emotions by referencing aspects of their identity in the personal-essential condition compared to all other story types. Still, between the ages of 4 and 7, children demonstrated increasing awareness of the limits to legitimate disobedience. That is, there was a significant age-related increase in predicting feel good compliance with moral rules that required the character to inhibit desires essential to his or her sense of self. By building bridges between theory of mind, social domain theory, and essentialism, this study reveals that it is not only possible to examine young children's knowledge about the importance of personal jurisdiction for self and identity but also that it is a rich area for continued exploration.

References

Arsenio, W. (1988). Children's conceptions of the situational affective consequences of sociomoral events. *Child Development*, *59*, 1611–1622.

Arsenio, W. F., Gold, J., & Adams, E. (2006). Children's conceptions and displays of moral emotions. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 581–609). Mahwah, NJ: Erlbaum.

Arsenio, W., & Lover, A. (1995). Children's conceptions of sociomoral affect: Happy victimizers, mixed emotions, and other expectancies. In M. Killen & D. Hart (Eds.), *Morality in everyday life: Developmental perspectives* (pp. 87–128). Cambridge, UK: Cambridge University Press.

Atance, C., & O'Neill, D. K. (2005). The emergence of episodic future thinking in humans. *Learning and Motivation*, *36*, 126–144.

Erikson, E. H. (1968). *Identity: Youth and crisis*. Oxford, UK: Norton.

Gelman, S. (2003). *The essential child: Origins of essentialism in everyday thought*. Oxford, UK: Oxford University Press.

Gelman, S. A. (2004). Learning words for kinds: Generic noun phrases in acquisition. In D. G. Hall & S. R. Waxman (Eds.), *Weaving a lexicon* (pp. 445–484). Cambridge, MA: MIT Press.

Gelman, S. A., & Heyman, G. D. (1999). Carrot-eaters and creature-believers: The effects of lexicalization on children's inferences about social categories. *Psychological Science*, *10*, 489–493.

Grusec, J. E., & Goodnow, J. J. (1994). Impact of parental discipline methods on the child's internalization of values: A reconceptualization of current points of view. *Developmental Psychology*, *30*, 4–19.

Harter, S. (1998). The development of self-representations. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology* (Vol. 3, 5th ed., pp. 553–618). New York: Wiley.

Harter, S., & Whitesell, N. R. (1989). Developmental changes in children's understanding of single, multiple, and blended emotion concepts. In C. Saarni & P. L. Harris (Eds.), *Children's understanding of emotion* (pp. 81–116). Cambridge, UK: Cambridge University Press.

Hasebe, Y., Nucci, L., & Nucci, M. S. (2004). Parental control of the personal domain and adolescent symptoms of psychopathology: A cross-national study in the United States and Japan. *Child Development*, *75*, 815–828.

Heider, F. (1946). Attitudes and cognitive organization. *Journal of Psychology*, *21*, 107–112.

Helwig, C. C. (2006). The development of personal autonomy throughout cultures. *Cognitive Development*, *21*, 458–473.

Helwig, C. C., Arnold, M. L., Tan, D., & Boyd, D. (2007). Mainland Chinese and Canadian adolescents' judgments and reasoning about the fairness of democratic and other forms of government. *Cognitive Development*, *22*, 96–109.

Helwig, C., Yang, S., Nucci, L., Yun, K., & To, S. (2009). Parental control of the personal domain and adolescent symptoms of psychopathology in urban and rural China. Paper presented at the biennial meeting of the Society for Research in Child Development, Denver, CO.

Helwig, C. C., Zelazo, P. D., & Wilson, M. (2001). Children's judgments of psychological harm in normal and noncanonical situations. *Child Development*, *72*, 66–81.

Higgins, E. T. (1991). Development of self-regulatory and self-evaluative processes: Costs, benefits, and tradeoffs. In M. R. Gunnar & A. L. Sroufe (Eds.), *Self processes and development: The Minnesota symposia on child development* (Vol. 23, pp. 125–166). Hillsdale, NJ: Erlbaum.

Kalish, C. W., & Shiverick, S. M. (2004). Children's reasoning about norms and traits as motives for behavior. *Cognitive Development*, *19*, 401–416.

Keller, M., Lourenço, O., Malti, T., & Saalback, H. (2003). The multifaceted phenomenon of "happy victimizers": A cross-cultural comparison of moral emotions. *British Journal of Developmental Psychology*, *21*, 1–18.

Killen, M., & Smetana, J. G. (1999). Social interactions in preschool classrooms and the development of young children's conceptions of the personal. *Child Development*, *70*, 486–501.

Kochanska, G., & Aksan, N. (2006). Children's conscience and self-regulation. *Journal of Personality*, *74*, 1587–1617.

Kusserow, A. S. (2004). *American individualisms: Child rearing and social class in three neighborhoods*. New York: Palgrave Macmillan.

Lagattuta, K. H. (2005). When you shouldn't do what you want to do: Young children's understanding of desires, rules, and emotions. *Child Development*, *76*, 713–733.

Lagattuta, K. H. (2007). Thinking about the future because of the past: Young children's knowledge about the causes of worry and preventative decisions. *Child Development*, *78*, 1492–1509.

Lagattuta, K. H. (2008). Young children's understanding of the influence of thoughts on emotions in rule situations. *Developmental Science*, *11*, 809–818.

Lourenço, O. (1997). Children's attributions of moral emotions to victimizers: Some data, doubts and suggestions. *British Journal of Developmental Psychology*, *15*, 425–438.

Nucci, L. P. (1981). Conceptions of personal issues: A domain distinct from moral or societal concepts. *Child Development*, *52*, 114–121.

Nucci, L. P. (1994). Mothers' beliefs regarding the personal domain of children. In J. G. Smetana (Ed.), *Beliefs about parenting: Origins and developmental implications* (pp. 81–97). San Francisco: Jossey-Bass.

Nucci, L. P. (1996). Morality and personal freedom. In E. S. Reed, E. Turiel, & T. Brown (Eds.), *Values and knowledge* (pp. 41–60). Mahwah, NJ: Erlbaum.

Nucci, L. P. (2001). *Education in the moral domain*. New York: Cambridge University Press.

Nucci, L. P. (2006). Education for moral development. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 657–682). Mahwah, NJ: Erlbaum.

Nucci, L. P., Camino, C., & Sapiro, C. M. (1996). Social class effects on northeastern Brazilian children's conceptions of areas of personal choice and social regulation. *Child Development*, *67*, 1223–1242.

Nucci, L. P., & Smetana, J. G. (1996). Mothers' concept of young children's areas of personal freedom. *Child Development*, *67*, 1870–1886.

Nucci, L. P., & Turiel, E. (2000). The moral and the personal: Sources of social conflicts. In L. P. Nucci, G. B. Saxe, & E. Turiel (Eds.), *Culture, thought, and development* (pp. 115–137). Mahwah, NJ: Erlbaum.

Nucci, L. P., & Weber, E. K. (1995). Social interactions in the home and the development of young children's conceptions of the personal. *Child Development*, *66*, 1438–1452.

Nunner-Winkler, G., & Sodian, B. (1988). Children's understanding of moral emotions. *Child Development*, *59*, 1323–1338.

Piaget, J. (1962). Will and action. *Bulletin of the Menninger Clinic*, *26*, 138–145.

Ruble, D. N., & Frey, K. S. (1991). Changing patterns of comparative behavior as skills are acquired: A functional model of self-evaluation. In J. Suls & T. Wills (Eds.), *Social comparison: Contemporary theory and research* (pp. 79–113). Hillsdale, NJ: Erlbaum.

Shaw, L. A., & Wainryb, C. (2006). When victims don't cry: Children's understandings of victimization, compliance, and subversion. *Child Development*, *77*, 1050–1062.

Smetana, J. G. (2002). Culture, autonomy, and personal jurisdiction in adolescent-parent relationships. In R. V. Kail & H. W. Reese (Eds.), *Advances in child development and behavior* (Vol. 29, pp. 51–87). San Diego, CA: Academic Press.

Smetana, J. (2005). Adolescent-parent conflict: Resistance and subversion as developmental process. In L. Nucci (Ed.), *Conflict, contradiction, and contrarian elements in moral development and education* (pp. 69–92). Mahwah, NJ: Erlbaum.

Smetana, J. (2006). Social-cognitive domain theory: Consistencies and variations in children's moral and social judgments. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 119–154). Mahwah, NJ: Erlbaum.

Stein, N. L., & Levine, L. J. (1989). The causal organization of emotional knowledge: A developmental study. *Cognition and Emotion*, *3*, 343–378.

Thompson, R. A., Meyer, S., & McGinley, M. (2006). Understanding values in relationships: The development of conscience. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 267–297). Mahwah, NJ: Erlbaum.

Tisak, M. S. (1993). Preschool children's judgments of moral and personal events involving physical harm and property damage. *Merrill-Palmer Quarterly, 39*, 375–390.

Turiel, E. (2002). *The culture of morality: Social development, context, and conflict*. New York: Cambridge University Press.

Wellman, H. M. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.

Wellman, H. M., & Banerjee, M. (1991). Mind and emotion: Children's understanding of the emotional consequences of beliefs and desires. *British Journal of Developmental Psychology, 9*, 191–214.

Wellman, H. M., & Lagattuta, K. H. (2000). Developing understandings of mind. In S. Baron-Cohen, H. Tager-Flusberg, & D. Cohen (Eds.), *Understanding other minds: Perspectives from developmental cognitive neuroscience* (2nd ed., pp. 21–49). New York: Oxford University Press.

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development, 75*, 523–541.

Wellman, H. M., & Miller, J. G. (2008). Including deontic reasoning as fundamental to theory of mind. *Human Development, 51*, 105–135.

Werner, H. & Kaplan, B. (1963). *Symbol formation*. Oxford, UK: Wiley.

Zelazo, P. D., & Mueller, U. (2002). Executive function in typical and atypical development. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (pp. 445–469). Malden, MA: Blackwell.

Child Development, March/April 2010, Volume 81, Number 2, Pages 636–651

Developmental Patterns in Decision-Making Autonomy Across Middle Childhood and Adolescence: European American Parents' Perspectives

Laura Wray-Lake, Ann C. Crouter, and Susan M. McHale
The Pennsylvania State University

Longitudinal patterns in parents' reports of youth decision-making autonomy from ages 9 to 20 were examined in a study of 201 European American families with 2 offspring. Multilevel modeling analyses revealed that decision-making autonomy increased gradually across middle childhood and adolescence before rising sharply in late adolescence. Social domain theory was supported by analyses of 8 decision types spanning prudential, conventional, personal, and multifaceted domains. Decision making was higher for girls, youth whom parents perceived as easier to supervise, and youth with better educated parents. Firstborns and secondborns had different age-related trajectories of decision-making autonomy. Findings shed light on the developmental trajectories and family processes associated with adolescents' fundamental task of gaining autonomy.

Developmental scholars have placed a high priority on understanding autonomy in youth (Baumrind, 1991; Collins, Gleason, & Sesma, 1997; Eccles et al., 1991; Holmbeck, Paikoff, & Brooks-Gunn, 1995; Smetana, Campione-Barr, & Daddis, 2004; Steinberg, 2001). Although autonomy can be defined in terms of behaviors, cognitions, or emotions, behavioral autonomy has been of particular interest (Collins et al., 1997; Hill & Holmbeck, 1987). Because it has long been acknowledged that becoming more autonomous is a central developmental task for adolescents (Erikson, 1968), most research and theorizing on autonomy has focused on adolescence. There is a paucity of longitudinal studies on autonomy across middle childhood and adolescence, however, and little is known about the nature of growth and change in behavioral autonomy as children become adolescents and move toward young adulthood.

The first goal of this study was to chart the trajectory of decision making—an indicator of behavioral autonomy—from middle childhood through adolescence using longitudinal data spanning ages 9–20 based on European American parents' reports. Everyday decisions about youth activities cut across a variety of social domains, and social domain theory predicts that autonomy levels will differ depending on the domain in question (Smetana, 1995; Turiel, 1998). Thus, a second goal was to examine the trajectories of eight decision-making items representing four domains: personal, social-conventional, prudential, and multifaceted. Based on an ecological perspective, the third goal was to examine correlates of decision-making trajectories including child characteristics (e.g., birth order, gender, openness to supervision) and family socioeconomic background.

Development of Decision-Making Autonomy

Behavioral autonomy refers to youth's freedom to regulate their own behavior as part of the process of developing independence and self-guided action (Collins et al., 1997; Feldman & Wood, 1994). Family decision making, a type of behavioral autonomy, refers to the everyday practices of making choices about youth activities, a process within which youth assert and parents negotiate control

---

An earlier version of this article was presented as a poster at the biennial meeting for the Society for Research on Child Development (2007, March) in Boston, Massachusetts.

We thank Megan Baril, Matt Bumpus, Kelly Davis, Doug Granger, Heather Helms, Julia Jackson-Newsom, Marni Kan, Wayne Osgood, Lilly Shanahan, Cindy Shearer, Kimberly Updegraff, and Shawn Whiteman for their help in conducting this study and the participating families for their time and insights about their family lives. This work was funded by a grant from the National Institute of Child Health and Human Development, RO1-HD32336. Ann C. Crouter and Susan M. McHale were co-principal investigators.

Correspondence concerning this article should be directed to Laura Wray-Lake, Department of Human Development and Family Studies, S-110 Henderson Building, The Pennsylvania State University, University Park, PA 16802. Electronic mail may be sent to ldw134@psu.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8102-0017

(Collins et al., 1997). Youth autonomy is minimal when parents make unilateral decisions for youth, moderate when parents and youth make joint decisions, and complete when youth make decisions unilaterally (Dornbusch et al., 1985; Lamborn, Dornbusch, & Steinberg, 1996). The development of behavioral autonomy has important implications: Both parent-unilateral and adolescent-unilateral decision making in adolescence have been linked to negative adjustment, whereas joint decision making is often seen as optimal for developmental and family functioning (e.g., Lamborn et al., 1996; Steinberg, Elmen, & Mounts, 1989). The level of decision making considered optimal, however, likely depends on what is normative for youth's age and developmental stage, as well as child and context characteristics.

Cross-sectional and longitudinal studies of adolescents have documented that family decision making becomes more adolescent driven and less parent unilateral with age (Dornbusch, Ritter, Mont-Reynaud, & Chen, 1990; Dornbusch et al., 1985; Eccles et al., 1991; Gutman & Eccles, 2007; Smetana et al., 2004). Using cross-sectional data, Dornbusch et al. (1985; Dornbusch et al., 1990) found that age was associated with more youth-unilateral and less parent-unilateral decision making. Joint decision making was highest during middle adolescence and lower in early and late adolescence (Dornbusch et al., 1985), suggesting that joint decision making is a transition between low and high decision-making autonomy. In other words, these studies suggest that the development of decision-making autonomy sequentially progresses through parent-unilateral, joint, and adolescent-unilateral decision making across adolescence.

Taking a more comprehensive longitudinal approach than previous studies, Gutman and Eccles (2007) examined decision-making autonomy using four waves of data spanning 8 years from the seventh grade to the year after high school. They found a quadratic pattern: Decision making increased across ages 13–19 and rose sharply between ages 15 and 17. These findings shed light on decision-making trajectories and suggest that steep increases occur in late adolescence, but the trajectory was estimated using only a single item reported by adolescents. Research into *autonomy expectations*, which are moderately correlated with behavioral autonomy, also provides insight into the timing of increase in autonomy across adolescence: Daddis and Smetana (2005) found that, on average, mothers expected adolescents to have autonomy for personal issues (i.e., decisions of individual preference) at ages 14–15 and to have autonomy for prudential issues (i.e., decisions with more negative consequences) at ages 16–17. In other longitudinal work, Smetana et al. (2004) studied an African American adolescent sample at three occasions over 5 years between ages 13 and 17 and reported a linear increase in decision-making autonomy. This finding was qualified by the social domain in which decisions were made, an issue discussed below. With only three occasions of measurement, however, complex longitudinal patterns could not be uncovered.

Longitudinal work on decision making has focused on adolescence, and little is known about the nature of autonomy in middle childhood. Decision-making processes do not always follow a simple linear progression across childhood and adolescence (Jacobs & Klaczynski, 2002). Charting how and when decision-making autonomy increases from middle childhood through adolescence, as our study does for a sample of European American families, fills a lacuna in the documentation of longitudinal developmental processes. Based on previous work (Daddis & Smetana, 2005; Gutman & Eccles, 2007), and using parents' reports, we expected a gradual increase in decision-making autonomy across ages 9–15, followed by a steep rise starting between ages 15 and 17.

Domain-Specific Decision Making

Social domain theory (Nucci, 2001; Smetana, 1995; Smetana, Crean, & Campione-Barr, 2005; Turiel, 1998) suggests that youth may achieve different levels of decision-making autonomy depending on the domain under which the decision falls. Autonomy is expected earliest in *personal domains* such as appearance (e.g., hairstyle, clothing) where decisions only impact the individual and pertain to private aspects of life (Daddis & Smetana, 2005; Feldman & Wood, 1994; Nucci, 2001; Smetana et al., 2004). Decisions falling under the *social-conventional domain* are those for which cultural norms within the family, community, or society serve as guidelines (e.g., manners, responsibility for household chores). Due to the possible negative consequences of violating such norms, lower autonomy is expected in the conventional as compared to the personal domain. *Prudential issues* are those that have potential negative consequences for health and safety, and thus the lowest levels of autonomy are expected in this domain (Smetana et al., 2004).

Multifaceted decisions reflect overlapping domains and typically traverse the personal and conventional or prudential domains (Smetana, 2000). For example, an adolescent's choice of friends is often seen as a personal issue by the adolescent but as a prudential issue by parents. Thus, multifaceted decisions may be a source of conflict for parents and adolescents (Smetana et al., 2005). Predictions about multifaceted decisions are least clear as there is little theorizing regarding the expected level of autonomy in relation to other domains (Baumrind, 2005). However, Smetana et al. (2004) predicted and found that adolescents had higher autonomy for multifaceted decisions than for those in conventional or prudential domains, as multifaceted decisions tend to be made jointly by parents and adolescents.

In the current study, youth's decision-making autonomy was measured by asking parents to report on youth decisions in eight areas: appearance, choosing activities, money, social life, bedtime/curfew, chores, homework/schoolwork, and health. Appearance lies in the personal domain, chores in the conventional, and health in the prudential domain; the remaining decision types are multifaceted. Even though these items are more generally worded than in previous research in social domain theory, longitudinal trajectories by domain represent a contribution to this literature. Based on theory and previous findings, we expected that parents' reports of youth autonomy levels would be highest for decisions about appearance, followed by the multifaceted decisions of choosing activities, homework/schoolwork, money, social life, and curfew. Autonomy levels should be lower for conventional decisions about chores than for personal or multifaceted decisions. Lowest autonomy levels were expected for (prudential) health decisions.

Along with testing predictions regarding level of differences across decision domains, we also examined differences in developmental trajectories by decision type. Smetana et al. (2004) found that both adolescents' and mothers' reports of autonomy in decisions across personal, conventional, prudential, and multifaceted domains increased gradually across ages 13–17, with one exception: Mothers' reports of adolescent autonomy in conventional decisions did not change. Our study examined decision-making autonomy over a longer time frame, and thus increases were expected across all domains. Based on Daddis and Smetana's (2005) finding that parents had expectations for earlier youth autonomy in personal decisions than in prudential ones, it is plausible that autonomy in conventional and prudential decisions may increase at a slower rate across adolescence. An alternative hypothesis, however, is that domains with lower autonomy levels in middle childhood and early adolescence would show steeper increases because they begin at a lower baseline and thus have farther to go before reaching full autonomy. Given the lack of previous research and theorizing, our examination of trajectory differences by domain was exploratory.

Correlates of Decision-Making Trajectories

From an ecological perspective, individual characteristics and family context help to shape youth development. As we elaborate, certain individual characteristics, such as gender and openness to supervision, may elicit more autonomy than others. In addition, two children within the same family may have different opportunities for decision-making autonomy based on their position in the family. Socioeconomic status (SES) may foster or hinder children's autonomy.

Gender. Findings on the role of gender in adolescent decision making are mixed. Some literature documents more autonomous decision making for boys (Dornbusch et al., 1990; Dowdy & Kliewer, 1998), and parents' autonomy expectations have also been found to be earlier for boys than for girls (Daddis & Smetana, 2005). In contrast, other literature suggests that girls experience greater decision-making autonomy than boys (Bumpus et al., 2001; Flanagan, 1990), and still other work finds no gender differences (Fuligni & Eccles, 1993; Peterson, Bush, & Supple, 1999; Smetana, 2000; Smetana et al., 2004). Gutman and Eccles (2007) found no main effects for gender but reported that girls and boys had different trajectories in decision-making autonomy: Girls' autonomy increased from ages 13 to 19, rising steeply between ages 15 and 17, whereas boys' autonomy increased linearly from ages 13 to 17 and afterward were stable. We examined gender differences in autonomy and tested for an interaction between gender and age, but proposed no hypotheses, given prior mixed findings and lack of theory.

Openness to supervision. Openness to supervision is defined as a youth's willingness to disclose information to parents and cooperate in the monitoring process (Kerns, Aspelmeier, Gentzler, & Grabill, 2001), and has been linked to greater parental knowledge of youth's activities and more positive parent–child relationships (Crouter, Bumpus, Davis, & McHale, 2005; Kerns et al., 2001). We measured parents' reports of children's openness to supervision in middle childhood, giving us the opportunity to examine how parents' perceptions of their children predict their reports of youth's autonomy development through adolescence. Children are active agents in their own development, and children's personal characteristics can influence family dynamics (Crouter & Booth, 2003). Children's openness to supervision should afford them greater involvement in decision making within the family, as cooperation engenders trust (Maccoby & Martin, 1983). We expected that parents' perceptions of children's openness to supervision would predict reports of higher levels and faster increases in youth decision-making autonomy from middle childhood through adolescence.

Birth order. Family theories imply birth order differences in decision-making autonomy. First, a growing body of literature based in a family systems perspective (Cox & Paley, 1997) suggests that parents with two children tend to learn from experience with their firstborn and adjust their parenting practices toward, expectations for, and relationships with their secondborn child (Shanahan, McHale, Crouter, & Osgood, 2007; Whiteman, McHale, & Crouter, 2003, 2007). These studies have shown that when siblings are compared at the same age, parents have less conflict with, more warmth toward, and more knowledge about secondborn compared to firstborn children. Regarding autonomy, the learning from experience model suggests that parents may relax expectations and allow autonomy earlier for secondborns compared to firstborns. Second, differentiation theory predicts that siblings endeavor to establish unique niches within the family in order to minimize their competition and maximize parental investment of resources (Ansbacher & Ansbacher, 1956; McHale et al., 2006; Sulloway, 1996). Building on Adler's theory of individual psychology, Sulloway (1996) argued that later born children seek autonomy in order to set themselves apart from more conforming and adult-oriented firstborns.

Both the learning from experience hypothesis and sibling differentiation theory imply greater autonomy for secondborns than for firstborns. Yet two cross-sectional studies found higher autonomy for firstborns than for secondborns (Bumpus et al., 2001; Small, Eastman, & Cornelius, 1988). In the majority of cross-sectional designs, birth order and age are confounded, in that later born siblings are also chronologically younger than earlier born siblings. Only in cross-sectional studies with substantial variability in the ages of sibling pairs can birth order and age be disentangled, and this concern has not been a focus of prior research. In contrast, a longitudinal design comparing siblings from the same family at the same age eliminates the confounding of age and birth order that characterizes most cross-sectional studies of siblings. Our data include 201 firstborn and secondborn sibling pairs that entered the study at roughly ages 11 and 9, respectively; our longitudinal design allowed us to make within-family comparisons to understand birth-order differences and study developmental changes for individuals. To fully illustrate the advantages of within-family comparisons related to birth order, we tested birth order differences in decision making by siblings' chronological age *and* by occasion of measurement, comparing firstborn and secondborn siblings' decision-making autonomy at the same age (using data from different measurement occasions) and at the same points in time (when the siblings differed in age). Using these two approaches, our goal was to reconcile differences between theoretical perspectives predicting greater autonomy for secondborns with cross-sectional findings showing greater autonomy for firstborns.

Parents' education. We treated parents' education as a proxy for social class, testing competing hypotheses regarding social class and youth decision making. According to Lareau (2003), parents from lower SES backgrounds value "accomplishment of natural growth" and leave children to create their own stimulation, thus granting them more autonomy. In contrast, parents of higher SES backgrounds value "concerted cultivation" and structure children's social lives and activities, resulting in lower autonomy for these youth. Consistent with this model, Flanagan (1990) found that mothers with long-term unemployment and less education reported greater decision-making autonomy for youth (particularly daughters).

Other studies, in contrast, suggest that higher SES predicts greater adolescent autonomy (e.g., Dornbusch et al., 1990; Nucci, Camino, & Sapiro, 1996). These studies are grounded in a broad literature suggesting SES differences in parenting styles. Lower SES parents tend to rely on more imperatives and restrictions, lowering youth autonomy, compared to higher SES parents (cf. Parke & Buriel, 2006). Thus, two bodies of literature suggest that parents' education may be related to youth decision-making autonomy in different ways. We examined the association between parents' education and youth decision making to test these competing hypotheses.

Study Goals

In sum, this study addressed three goals. First, we examined the trajectory of global decision making from middle childhood through late adolescence using reports from European American parents. Second, we charted trajectories of eight decision-making items in four domains to test hypotheses derived from the social domain literature. Third, we tested gender, youth's openness to supervision, birth order, and parents' education as potential correlates of the development of decision-making autonomy from middle childhood through adolescence.

Method

Sample and Procedure

Data were drawn from seven waves (Waves 2, 3, and 6 through 10) of a 10-year longitudinal study exploring family relationships and gender development across middle childhood and adolescence. Waves 1, 4, and 5 could not be included because the decision-making measure was not available at these waves. Waves 2, 3, and 6 through 10 are referred to here as Occasions 1 through 7. Families were recruited through letters sent home with fourth- and fifth-grade students in 16 school districts. Interested families returned self-addressed stamped postcards if they met three criteria: (a) a mother and father (either biological or adoptive) present in the household, (b) a firstborn child in the fourth or fifth grade, and (c) at least one sibling 1–4 years younger. Of the eligible families who returned postcards, over 90% agreed to participate in the study. Comparisons of participating families with U.S. Census data suggested that parents in the sample were slightly older and better educated than dual-earner families from the same counties.

Of the 203 families that originally agreed to participate, 2 families dropped out after the first wave and were excluded here. Thus, current analyses used 201 families. A team of interviewers conducted separate 2- to 3-hr home interviews with fathers, mothers, and the two target siblings in each family. Families received a $100 honorarium in Waves 1, 2, and 3 and $200 thereafter. The last point of data collection for each family was the year after the firstborn sibling graduated from high school, which varied across families, and meant that the sample size dropped beginning at Occasion 4. Attrition was low: At Occasion 4, the last wave in which all youth participated, 95% of the original sample remained. Across the study, 9 families experienced divorce, 3 sets of parents separated, a parent's death affected 6 families (5 were fathers), and 2 adolescents had children of their own. Of the 177 individuals who reached the ages of 19 or 20 by the end of the study (mostly firstborns), the majority were attending a 4-year college (66%). In addition, 18% were employed full- or part-time, 12% attended a 2-year or vocational school, and 4% had joined the military. Finally, 24% lived at home with their parents.

The sample was composed of White, working- and middle-class families residing in small cities and rural areas. With the exception of 2 adopted Asian American children, all families were European American, which was representative of the region. At Occasion 1, average family income was $63,355 (*SD* = 31,472), slightly wealthier than other families in the area. Mothers and fathers averaged 14.63 (*SD* = 2.11) and 14.72 (*SD* = 2.40) years of education, respectively, representing vocational training/some college. Firstborn children (103 girls and 98 boys) were 11.83 years of age on average (*SD* = 0.55), and secondborns (100 girls and 101 boys) averaged 9.22 years of age (*SD* = 0.93).

Measures

Decision making. The adolescent decision-making questionnaire was adapted from Dornbusch et al. (1985), and domains were taken from Smetana (1988). Only parent reports were collected across the entire study period. The decision-making measure consisted of eight items. Mothers and fathers were asked to "think about how decisions have been made during the past year in different areas of your child's life." They circled the number corresponding to the family member(s) who usually made decisions in eight areas: chores, appearance, homework/schoolwork, social life, bedtime/curfew, health, choosing activities, and money. Appearance was classified under the personal domain, chores under the conventional domain, and health under the prudential domain; all other decision types were classified as multifaceted. Original response options ranged from 1 to 9, representing the following: (1) child alone, (2) mother, (3) father, (4) both parents, (5) father and child, (6) mother and child, (7) parents and child, (8) other person(s), and (9) nobody. Items were recoded to a 3-point scale to indicate decisions made by: (1) one or both parents, (2) child and one

Table 1

Means, Standard Deviations, and Correlations for Global Decision-Making Index for Each Measurement Occasion and Continuous Predictors

		M (*SD*)		Correlations								
	N	Firstborns	Secondborns	1	2	3	4	5	6	7	8	9
1. DM 1	402	1.76 (0.27)	1.67 (0.25)	–	.54***	.37***	.40***	.32***	.39***	.22	.13†	.28***
2. DM 2	402	1.83 (0.24)	1.74 (0.23)	.59***	–	.49***	.49***	.42***	.42***	.40*	.19**	.18†
3. DM 3	387	2.02 (0.25)	1.89 (0.22)	.38***	.43***	–	.64***	.52***	.43***	.55**	.14*	.19**
4. DM 4	380	2.10 (0.26)	1.94 (0.24)	.39***	.40***	.59***	–	.61***	.57***	.44*	.11	.14†
5. DM 5	379	2.31 (0.30)	2.00 (0.26)	.25***	.31***	.48***	.62***	–	.74***	.51**	.17*	.10
6. DM 6	231	2.45 (0.28)	2.04 (0.24)	.21*	.27**	.38***	.51***	.61***	–	.52**	.12	.22*
7. DM 7	53	2.49 (0.24)	2.09 (0.24)	.28	.09	.34†	.38†	.31	.60***	–	.20	−.14
8. Open to supervision	401	4.23 (0.47)	4.16 (0.51)	.11	.16*	.10	.04	−.03	.10	.05	–	.20**
9. Parents' education	402	14.67 (2.0)		.26***	.26***	.22**	.07	.03	−.01	−.23	.23***	–

Note. *N*s represent number of adolescents, including both siblings. Correlations for firstborns are below the diagonal and correlations for secondborns are above the diagonal. DM = decision making.
†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

or both parents, and (3) child alone. The 3-point scale captures the dimensions of parent-unilateral, joint, and adolescent-unilateral decision making used in the previous literature (Dornbusch et al., 1985). Autonomy is conceived of as a continuum from parent unilateral to adolescent unilateral with joint decision making as the midpoint. A continuous scale was considered most appropriate for illustrating developmental patterns. Responses of "nobody" or "other person(s)" were rare (.01% and .001% of all responses across waves) and were coded as missing.

Collapsed across occasions and reports for each child, mother and father reports were moderately correlated ($r = .50$, $p < .001$). Thus, mother and father reports of decision-making autonomy were averaged. In cases where one parent's report was missing, data from the available parent were used. The decision to average parents' reports was supported by additional analyses (available upon request) showing the same longitudinal pattern for mothers and fathers as well as high percent agreement on reports for each child and each item. Across occasions, mothers ($M = 1.94$) reported higher decision-making autonomy for secondborns than did fathers ($M = 1.77$), $t(200) = 9.75$, $p < .001$; the difference between mother ($M = 2.05$) and father ($M = 2.02$) reports of decision making for firstborns was marginal, $t(200) = 1.80$, $p < .10$.

A global decision-making index was created by averaging all items. In addition, each item was examined separately. Creation of the global decision-making index was supported by satisfactory Cronbach's alpha coefficients for each offspring at each wave (α range = .74–.83, $M = 0.79$). Higher item and index means on decision-making items reflected greater autonomy in decision making. Means for the decision-making index appear in Table 1.

Openness to supervision. At Occasion 2, mothers and fathers rated offspring's openness to supervision using 12 items adapted from the Child Check-In Scale (Kerns et al., 2001), for example, "My child returns periodically throughout the day to check in." Response options ranged from *not at all* (1) to *very often* (5). Mother and father reports were moderately correlated for firstborns ($r = .36$, $p < .001$) and secondborns ($r = .40$, $p < .001$), and were averaged. Higher scores reflected greater openness to supervision. Cronbach's alphas were satisfactory (αs = .89 and .92 for firstborn and secondborn offspring, respectively). Firstborns were perceived as easier to supervise ($M = 4.23$), on average, than secondborns ($M = 4.16$), $t(199) = -2.31$, $p < .05$. Firstborn girls were perceived as easier to supervise ($M = 4.30$) than firstborn boys ($M = 4.15$), $t(199)=2.26$, $p < .05$, but there were no differences in openness to supervision for secondborn girls ($M = 4.21$) versus boys ($M = 4.10$), $t(198) = 1.53$, *ns*.

Parents' education. We treated parents' education in years as a correlate of decision making. Both mothers' and fathers' levels of education were highly stable over time. The average correlation across all occasions was .96 for mothers and .84 for fathers. Mothers' and fathers' education was moderately correlated ($r = .45$, on average, across waves). Thus, we used the average of both parents' education at Occasion 1 ($M = 14.67$, $SD = 2.00$).

Table 1 shows correlations between the global decision-making index and continuous predictors. Parents' reports of decision-making autonomy were highly correlated across waves; Occasion 7 correlations were less strong, probably due to the smaller sample size that was an artifact of the study's design (i.e., following families until the year after the firstborn's high school graduation). Parents' education was more highly correlated with decision-making autonomy in earlier waves. Correlations between openness to supervision and decision-making autonomy were low.

Results

Analytic Plan

Multilevel modeling analyses were conducted using SAS PROC MIXED 9.1. Unconditional polynomial growth models were estimated for the global decision-making index and for each of the eight decision-making items. Correlates of gender, children's openness to supervision, birth order, and parents' education were examined in each model.

Our nested data design called for three levels. The *Level 1*, or within-person, model captured predictors of intraindividual change in decision making over time, and time (defined as age of the child) was the unit of analysis. Time was centered at age 10, the earliest age at which both siblings were surveyed. *Level 2* captured the between-sibling (or within-family) effect, and thus the unit of analysis was individual siblings (i.e., firstborns, secondborns). Effects estimated at Level 2 were gender (girl = 0, boy = 1), birth order (firstborn = 0, secondborn = 1), openness to supervision (mean centered), and interaction terms involving those variables. *Level 3* included parents' education (mean centered) as a between-family effect (shared by both offspring).

We estimated a series of models, beginning with unconditional growth models, to describe change in global decision making and decision-making types. Prior to adding predictors to the global decision-making index model, intraclass correlation coefficients revealed substantial variance at each level: Seventy-four percent of the variance at Level 1 (within person) was unshared, and 48% of variance at Level 2 (between siblings) was unshared. Restricted maximum likelihood estimates were used for reporting model parameters and assessing the significance of random effects. As each correlate was added to the models, interactions were tested with fixed effects of time, birth order, and gender. Nonsignificant interactions were pruned before testing additional effects.

Pattern of Change for Global Decision-Making Index

The overall pattern of change for the global decision-making index was a positive cubic effect of age (see Table 2 and Figure 1). The best fitting unconditional growth curve called for a random linear term at Level 3 and a random intercept at Level 2; in other words, a linear slope was estimated for each family and an intercept was estimated for each child.

As Figure 1 illustrates, parents' reports of youth input into decision making increased gradually in middle childhood (age 9–11), remained relatively flat across early adolescence (age 11–13), and increased more steadily across middle adolescence (age 13–15). After age 15, the curve rose steeply, indicating that adolescents began to experience greater autonomy in decision making. Thus, the overall pattern was a gradual yet steady increase in decision making across middle childhood through middle adolescence, followed by a rapid increase in

Table 2
Unconditional Growth Curve Model for Global Decision-Making Index

	Global decision-making index		
Parameter	Estimate	Standard error	*t* value
Fixed effects			
Intercept (age 10)	1.73	.02	111.24***
Linear	0.05	.005	9.60***
Quadratic	−0.01	.002	−6.38***
Cubic	0.001	.0001	11.42***
Level 2 variance			
Intercept	0.003	.001	3.29***
Level 3 variance			
Intercept	0.04	.005	7.62***
Intercept-linear	−0.003	.001	−5.08***
Linear	0.0007	.0001	6.76***
Residual	0.03	.001	28.25***
Model fit			
REML deviance	−813		
AIC	−803		
BIC	−786		

Note. REML = restricted maximum likelihood; AIC = Akaike's information criterion; BIC = Bayesian information criterion.
***$p < .001$.

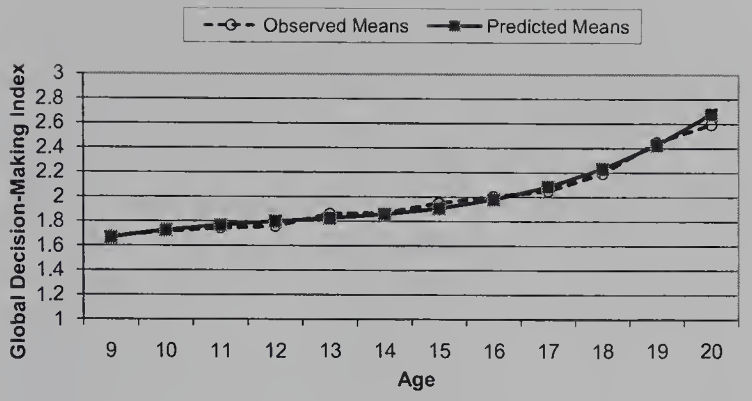


Figure 1. Average trajectory and observed means for global decision-making index by age.

decision-making autonomy in middle to late adolescence. The decision-making scale ranged from 1 to 3, with a score of 2 indicating joint parent–offspring decision making. At age 18 and after, mean levels of decision making approached but did not reach full autonomy, suggesting that joint decision making between parents and adolescents remained normative in most domains.

Mean Differences for Decision-Making Items

After specifying the growth curve for the decision-making index, we turned to item-level analyses. As described above, social domain theory posits different levels of decision-making autonomy depending on the domain. To test for overall level differences, we averaged each item across all occasions, and conducted repeated measures analyses of variance (ANOVAs) separately on parents' reports for firstborns and secondborns, repeating on the eight decision-making domains (means shown in Figure 2). Using contrast statements, means for each domain were contrasted with every other domain. Given the 28 comparisons for each child, we used a stringent error rate of $p < .001$ to evaluate significant differences.

For firstborns, all means were significantly different from the others with a few exceptions: Decision-making means for schoolwork ($M = 2.21$), activities ($M = 2.20$), and money ($M = 2.14$) were not significantly different, and means for money and social life ($M = 2.13$) were not different (see Figure 2). A similar pattern emerged for secondborns, yet fewer differences were found: Decision making for money ($M = 2.06$) was not significantly different from means for appearance ($M = 2.07$); schoolwork ($M = 1.96$), activities ($M = 1.98$), and social life ($M = 1.97$) were not different; and health ($M = 1.64$) was not different from curfew ($M = 1.68$).

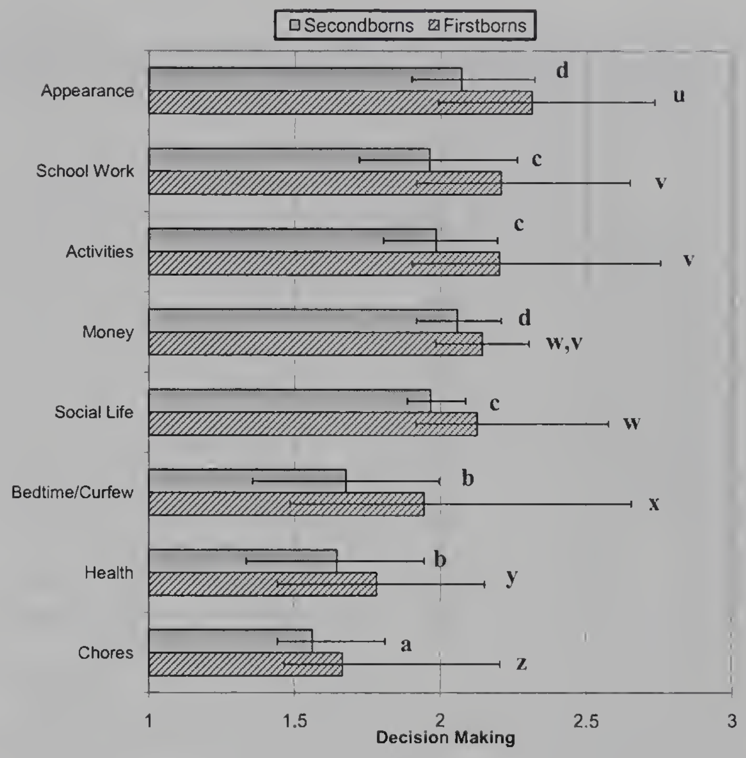


Figure 2. Across-time decision-making item means for firstborn and secondborns.

Note. Means represent averages pooled across seven occasions. Bars represent range of change over time for each item. Secondborn differences (shaded bars) of $p < .05$ are denoted by a, b, c, and d. Firstborn differences (striped bars) of $p < .001$ are denoted by u, v, w, x, y, and z.

In accordance with the predictions of social domain theory (Smetana, 1995; Smetana et al., 2004), the personal domain of appearance had the highest mean level of autonomy compared to all other domains for firstborns ($M = 2.31$). As expected, low levels of autonomy were found for the prudential domain of health for both children (see Figure 2). Unexpectedly, however, autonomy was lower for chores than for health for both firstborns and secondborns. As hypothesized, means for multifaceted decisions generally fell in between highest means of the personal domain and lowest means of the conventional and prudential domains. For secondborns, however, decision-making autonomy for money was as high as the personal domain of appearance, and autonomy for curfew was as low as the prudential domain of health.

Growth Curves by Decision-Making Type

We then examined growth curves by decision type to determine whether developmental trajectories differed by decision-making domain. Fixed and random effects for final models are shown in Table 3. Figure 3 depicts patterns in predicted means for parents' reports of decision-making

Table 3
Fixed and Random Effects for Unconditional Growth Curve Models by Item

	Fixed effects			
	Intercept	Linear	Quadratic	Cubic
Appearance	1.93 (.02)***	.04 (.01)***	−.004 (.003)	.001 (.0002)***
Schoolwork	1.82 (.02)***	.08 (.01)***	−.02 (.003)**	.002 (.0003)***
Activities	1.86 (.02)***	.04 (.01)***	−.01 (.003)**	.002 (.0002)***
Money	1.94 (.02)***	.04 (.003)***	–	–
Social life	1.93 (.02)***	.03 (.01)**	−.02 (.002)***	.002 (.0002)***
Bedtime/Curfew	1.45 (.02)***	.07 (.01)***	−.01 (.003)	.002 (.0003)**
Health	1.37 (.02)***	.08 (.003)***	–	–
Chores	1.48 (.02)***	.03 (.01)**	−.01 (.003)*	.001 (.0002)***

	Random effects				
	Level 2	Level 3			
	Intercept	Intercept	Linear	Quadratic	Residual
Appearance	.01 (.003)***	.05 (.01)***	.002 (.0003)***	–	.10 (.003)***
Schoolwork	.03 (.01)***	.05 (.01)***	.001 (.0002)***	–	.11 (.004)***
Activities	.002 (.002)	.05 (.01)***	–	–	.10 (.004)***
Money	.01 (.003)*	.08 (.01)***	.002 (.001)***	–	.16 (.01)***
Social life	.0004 (.002)	.06 (.01)***	.003 (.001)**	.00003 (.00001)***	.07 (.003)***
Bedtime/curfew	.01 (.003)*	.05 (.01)***	.003 (.001)*	.00004 (.00002)*	.12 (.01)***
Health	.002 (.002)	.04 (.01)***	.001 (.0002)***	–	.10 (.003)***
Chores	.003 (.003)	.06 (.01)***	.001 (.0002)***	–	.11 (.004)***

Note. Standard errors appear in parentheses.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

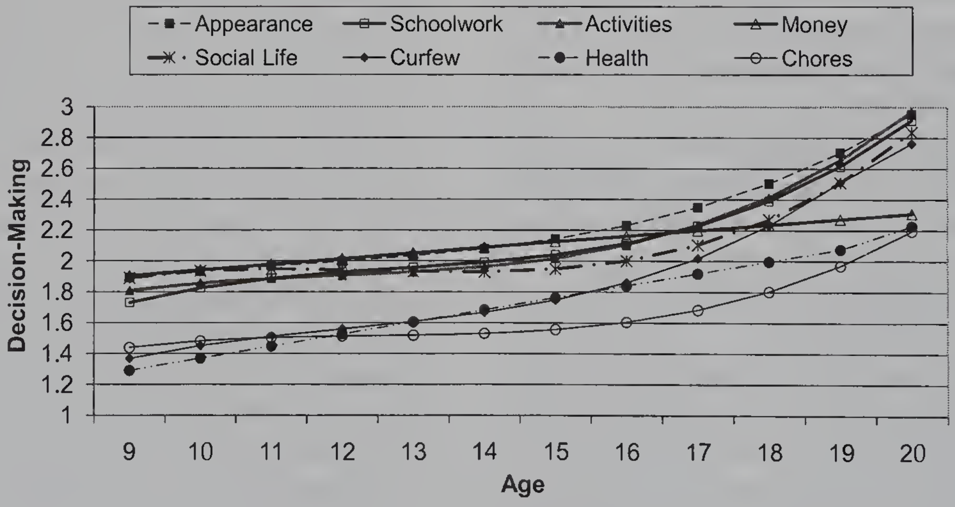


Figure 3. Developmental trajectories of decision making by item. *Note*. Predicted means from growth curve models are displayed.

items by age. Unconditional growth curves revealed that six of eight decision-making items followed a cubic pattern, as shown by significant fixed effects of age up to the cubic polynomial. Consistent with the global decision-making index, these curves were relatively stable in middle childhood, increased gradually during early and middle adolescence, and increased sharply during late adolescence. Only decisions made about money and health followed a different, linear pattern, with parents' reports of autonomy in decisions of these types increasing gradually and steadily from ages 9 to 20.

Item-level trajectories elaborated on the findings of average level differences reported above. Items seemed to form two groups reflecting higher and lower autonomy in middle childhood and again in late adolescence (see Figure 3); these observations were supported by repeated measures ANOVAs of mean differences by item at the beginning (ages 9–11) and end (ages 17–20) of the trajectories. In middle childhood, means for chores, health, and curfew fell close to parent-unilateral decision making ($M = 1.43$, $SD = 0.29$), reflecting low autonomy. The remaining items reflected high autonomy in middle childhood, approaching joint decision making ($M = 1.87$, $SD = 0.27$). Decision types that started high tended to remain higher in autonomy across time, and types that started low tended to remain lower in autonomy across time, with two exceptions. In late adolescence, decisions about money were still made jointly. Thus, in late adolescence, chores, money, and health reflected joint decision making on average ($M = 2.01$, $SD = 0.28$). Decisions about curfew started low in autonomy ($M = 1.37$ at age 9) but showed high autonomy in late adolescence ($M = 2.70$ at age 20), exhibiting substantial change over time. Overall, decisions regarding curfew, social life, activities, schoolwork, and appearance, when averaged in late adolescence, fell between joint decision making and adolescent-unilateral decision making ($M = 2.36$, $SD = 0.38$).

Table 4
Global Decision-Making Index Model with Correlates

	Global decision-making index		
Parameter	Estimate	Standard error	*t* value
Fixed effects of age			
Intercept	1.58	.05	32.05***
Linear	0.17	.03	5.03***
Quadratic	−0.04	.007	−5.32***
Cubic	0.003	.0004	7.61***
Child characteristics			
Birth order (0 = FB, 1 = SB)	0.16	.05	3.37***
Birth Order × Linear Age	−0.13	.03	−3.81***
Birth Order × Quadratic Age	0.03	.007	4.86***
Birth Order × Cubic Age	−0.003	.0005	−5.56***
Gender (0 = girl, 1 = boy)	−0.04	.01	−3.21**
Openness to supervision	0.06	.02	3.45***
Parent characteristics			
Parents' education	0.01	.006	2.60**
Model fit			
REML deviance	−674		
AIC	−664		
BIC	−648		

Note. FB = firstborn; SB = secondborn; REML = restricted maximum likelihood; AIC = Akaike's information criterion; BIC = Bayesian information criterion.
$^{**}p < .01$. $^{***}p < .001$.

Correlates of Decision Making

To address our third goal, we examined youth characteristics and parents' education as correlates of parents' reports of global decision making. The final global decision-making model (see Table 4) explained 31% of within-person (Level 1) variance and 33% of between-sibling (Level 2) variance.

There was a significant main effect of gender on global decision making: Parents reported more decision-making autonomy for girls than for boys, and girls were higher, on average, by a factor of .04 units. Openness to supervision was positively associated with parent reports of global decision-making autonomy: For each ease of supervision unit above average, decision-making autonomy was higher on average by .06 units. Parents' education was significantly and positively related to global decision-making autonomy. For each year of parents' education over the sample average, decision making was higher on average by .01 units. The effects of gender, openness to supervision, and parents' education were consistent over time and did not interact with other predictors.

Turning to birth order, a significant main effect of birth order indicated that parents reported greater autonomy for secondborns than firstborns at age 10 ($\beta = .16$, $p < .001$). This birth order effect was qualified by interactions with linear, quadratic and cubic age effects (see Table 4). These effects indicated a significant statistical test of different age-related trajectories for firstborn and secondborn siblings: Firstborns showed a clear cubic pattern, like the average decision-making trajectory described above. In contrast, secondborns increased gradually and linearly in decision-making autonomy with age. Although secondborns had higher decision-making autonomy than firstborns at age 10, this birth order difference lessened with age (see Figure 4a). Reestimating the model with the intercept at ages 13 ($\beta = .01$, *ns*), 15 ($\beta = .04$, $p < .01$), and 18 ($\beta = -.02$, *ns*) confirmed the attenuated effect of birth order on decision-making autonomy with age.

To reconcile our findings with the extant literature, we contrasted the effect of birth order when siblings were *the same age* with the effect of birth order when parents were asked at *the same occasion* about each offspring's decision-making autonomy. To conduct this analysis, we treated occasion as the metric of time instead of age. After first fitting an unconditional growth model to decision-making

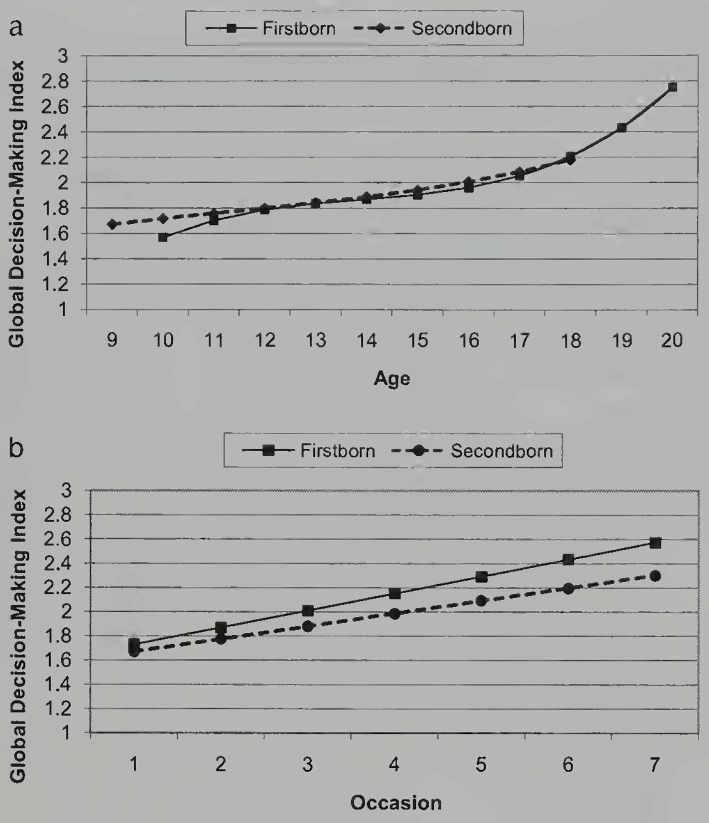


Figure 4. Birth Order × Age (a) and Birth Order × Occasion (b) interaction for global decision-making index.

autonomy by occasion, we entered birth order and the birth order by occasion interaction. Occasion was centered at the first occasion, which was most comparable to centering (age 10) in the age-based model. The best-fitting unconditional growth curve was a fixed linear effect, with a random Level 2 intercept and random linear slope at Level 3.

In the occasion-based model, parents reported significantly higher decision-making autonomy for firstborns at Occasion 1 than for secondborns at the same occasion by a factor of .06 units. This birth order effect was qualified by an interaction with occasion (see Figure 4b). The interaction indicated that the difference between siblings, favoring firstborns, increased across measurement occasions. Thus, although secondborns had higher decision making than firstborns when measured at age 10 and this effect became smaller with time, firstborns had higher decision making than secondborns at the first occasion, and this effect became larger with time. Firstborns reached the oldest ages in the study (19 and 20; the year after high school graduation) and showed steep increases in autonomy at these ages (see Figure 4a), whereas secondborns were not surveyed at these ages. This difference helps to explain why firstborns had greater decision-making autonomy on average across the study compared to secondborns (Figures 2 and 4b) and why the overall pattern looks more similar to the firstborn trajectory (Figure 1). Yet during middle childhood, parents reported more autonomy for secondborns than firstborns.

Discussion

This study provides a comprehensive longitudinal examination of youth decision-making autonomy in European American families from the perspective of parents. Developing autonomy is a fundamental task of middle childhood and adolescence (Erikson, 1968), and part of the developmental process involves mothers and fathers exercising authority and negotiating control with children (e.g., Smetana, 1995). Thus, mother and father reports of youth decision-making autonomy are important sources for information about family processes. Rather than finding one normative developmental pattern, our study charted variability in the development of decision-making autonomy that depended on decision domain as well as birth order. The findings from parents' reports of global and specific types of decision making enhance our understanding of the development of behavioral autonomy across middle childhood and adolescence and its correlates. In concluding, we summarize the developmental patterns, couch our findings in social domain theory, review the correlates of decision making, discuss limitations, and point to directions for future research.

Overall Developmental Patterns

Global decision-making autonomy increased gradually from middle childhood through adolescence, and showed a steep increase in late adolescence after age 15. This pattern adds specification to the past literature documenting increases in behavioral autonomy in adolescence (Dornbusch et al., 1985; Dornbusch et al., 1990; Gutman & Eccles, 2007; Smetana et al., 2004). For both older and younger siblings, the period spanning middle and late adolescence, from age 15 to 20, was characterized by higher autonomy compared to middle childhood and earlier adolescence.

Analyses of both global and domain-specific decision-making trajectories revealed that joint decision making was normative in our sample across middle childhood and adolescence. Five of the eight decision-making items showed joint decision making starting in middle childhood, and means remained at or below joint decision making for money, health, and chores between ages 17 and 20. The latter findings confirm that parents still have input into some offspring decisions in late adolescence. In middle childhood, joint decision making is not always a given, as our findings revealed that children, before adolescence, had little input in decisions about chores, curfew, and health. Many studies have suggested that joint decision making between parents and offspring has positive implications for adolescents' adjustment (e.g., Lamborn et al., 1996; Peterson et al., 1999; Steinberg et al., 1989). Moreover, behavioral autonomy is defined by independent decision making tempered with support from family and others (Collins et al., 1997; Hill & Holmbeck, 1987). Based on a person–environment fit perspective (Eccles et al., 1991), an ideal family context should provide levels of control and autonomy that entail not giving too much autonomy in middle childhood and not giving too little in late adolescence; though needing replication in other samples, the overall developmental trajectory found for parents' reports in our sample reflects such a pattern.

Social Domain Theory

The tenets of social domain theory were generally supported by our analyses of level differences as a function of decision domain. For both firstborn and secondborn siblings, decisions in the personal domain garnered the most autonomy, decisions in the prudential and conventional domains garnered the least, and multifaceted domains typically fell in the middle. Youth had lower autonomy in decisions about chores than about health, and developmental trajectories also showed that autonomy for decisions about chores was lower than for decisions about health across adolescence. These findings deviated from social domain theory's prediction of lowest autonomy levels for prudential decisions, yet autonomy for health decisions was lower than other domains, as predicted. Findings for the conventional domain of chores were consistent with the findings of Smetana et al. (2004) that autonomy levels were similar for prudential and conventional decisions, that mothers and adolescents tended to view conventional domains as parent decisions, and that mothers' reports of youth autonomy in conventional decisions (i.e., chores, how to talk to parents, language, and manners) remained low over time. Mean differences in decision types for secondborns were not as strong or as consistent with social domain theory as were findings for firstborns, likely because decision-making autonomy was lower, on average, across domains for secondborns.

Item-level developmental trajectories of decision-making autonomy showed variability in terms of growth patterns. Unlike the majority of decision types that followed a cubic pattern of change with age, autonomy for decisions about health and money increased gradually in a linear fashion. Health decisions fit with the conceptual argument that when families negotiate lower levels of adolescent autonomy on prudential decisions, autonomy in these types of decisions also increases at a slower rate across adolescence. The linear increase in autonomy for decisions about money—with means hovering around joint decision making across the study—represents a novel finding to be considered in future studies of social domain theory. Previous work has measured aspects of decisions about money more specifically and categorized them in the personal domain (e.g., Smetana et al., 2004), yet this distinct linear pattern was not consistent with patterns for either the multifaceted or personal domains. It is unclear whether our findings regarding money warrant a substantive conclusion or represent a measurement issue, as the general nature of our question may be responsible for this finding. Future research should consider differentiating the sources of adolescents' money (e.g., job earnings, allowance, gifts), as these sources may relate to families' decisions about autonomy with money.

Overall, similar developmental trajectories emerged for the majority of decision-making types, supporting the use of a global decision-making index. However, decisions about money and health followed linear trajectories, which diverged from the general pattern and thus represent novel findings. Indeed, this study provides the first findings of trajectory differences in decision domains across middle childhood and adolescence and informs future theorizing about domain-specific developmental changes in decision-making autonomy. Level differences in item means and trajectories were prominent, yet necessarily obscured, when decision making was examined as a global index. In short, the global and specific approaches convey important but distinct information about decision-making autonomy.

Correlates of Offspring Decision Making

Parents reported greater decision making for daughters, for youth initially viewed as easier to supervise, and for youth with better educated parents. Furthermore, different age-related trajectories in decision-making autonomy emerged for firstborn and secondborn siblings. Each of these findings is a novel addition to the literature.

Our results converged with those of several prior studies (Bumpus et al., 2001; Flanagan, 1990) in documenting that girls have more autonomy in decisions than do boys. Girls may have greater decision-making autonomy as a function of their relative maturity (Bumpus et al., 2001; Flanagan, 1990). Our results are inconsistent with studies that found no gender differences (e.g., Fuligni & Eccles, 1993; Smetana et al., 2004) or differences favoring boys (e.g., Dornbusch et al., 1990). Differences across studies may reflect sample variation in terms of families' gender role traditionality, and autonomy in global or domain-specific decisions may depend on gender socialization (e.g., Bumpus et al., 2001; Crouter, Whiteman, McHale, & Osgood, 2007).

As expected, openness to supervision in middle childhood predicted greater decision-making autonomy. Being perceived as easy to supervise may indicate to parents that youth can be trusted and are ready to handle more autonomy (Fuligni & Eccles, 1993; Kerns et al., 2001). This effect was moderate in effect size (Cohen's $d = .4$), and as openness to supervision was only measured once in middle childhood, its predictive power suggests that this characteristic matters for autonomy development.

Our finding of a positive association between parents' education and youth decision-making autonomy is consistent with some prior work (Dornbusch et al., 1990; Nucci et al., 1996) and with the link between social class and autonomy found in the broader parenting literature (cf. Parke & Buriel, 2006). Our findings did not support Lareau's (2003) perspective that parents of lower social status offer more autonomy to encourage ''natural growth'' in their children. Instead, our findings suggest that parents with less formal education may emphasize parental authority, perhaps to create a safer environment (Parke & Buriel, 2006), whereas more educated parents emphasize self-direction (Kohn, 1977; Nucci et al., 1996). Further research is needed to identify the mechanisms underlying this association.

Our longitudinal within-family design enabled us to reconcile discrepancies between theory and empirical findings regarding birth order and autonomy. The learning from experience perspective (e.g., Whiteman et al., 2003) and sibling differentiation theory (e.g., Sulloway, 1996) posits greater autonomy for secondborn compared to firstborn children. Empirical findings that older siblings have more decision-making autonomy than younger siblings have been cross-sectional (Bumpus et al., 2001; Small et al., 1988). Consistent with these results, when we used occasion of measurement as the time metric, firstborns had higher decision-making autonomy than secondborns. In contrast, by treating *age* as the time metric, that is, when siblings were compared at the same age, we found that secondborns had greater autonomy in decisions than firstborns, primarily in middle childhood. The latter finding provides some support for the idea that parents relax expectations for secondborns and that younger offspring seek autonomy in an effort to set themselves apart from older siblings. These birth order effects were particularly prominent in middle childhood and early adolescence, as all youth increased in decision-making autonomy over time. Firstborns tended to increase in decision-making autonomy in a cubic fashion much like the overall pattern whereas secondborns' autonomy increased gradually and linearly. Overall, these findings suggest that developmental patterns of decision-making autonomy may be dependent on the family context. The different trajectories may be in part explained by the later ages that firstborns, in contrast to secondborns, reached in our study. Further investigations in other samples are needed to replicate these within-family patterns.

Limitations and Directions for Future Research

This study highlights the merits of longitudinal, within-family designs for understanding how developmental patterns of autonomy unfold over time. An important next step for future research is extending the examination of autonomy further into young adulthood. Arnett (2000) reported that a key criterion for being considered an adult is making independent decisions, yet today's young adults are only ''semi-autonomous.'' Data are needed to examine the circumstances under which joint decision making versus complete autonomy becomes the norm in young adulthood. More data could also demonstrate whether the trajectory for secondborns conforms to firstborns' pattern in late adolescence or whether it remains distinct. Future research into decision-making autonomy may benefit from using measures with a larger range than our 3-point scale in order to capture developmental shifts with greater sensitivity.

Our findings supported and extended the tenets of social domain theory, yet several questions remain unaddressed. Unexpectedly, based on previous work (Smetana et al., 2004), the conventional domain of chores showed a steep rise in late adolescence rather than increasing gradually. Perhaps parents are more restrictive of autonomy in conventional domains like chores in middle childhood and earlier in adolescence, an idea needing further investigation. Though, as predicted, multifaceted decisions showed autonomy levels lower than those in the personal domain and higher than those in conventional and prudential domains, multifaceted decisions represent an ambiguous category that deserves further discussion among social domain theorists. For example, Baumrind (2005) suggests that multifaceted decisions should be classified based on the domain (e.g., conventional, prudential) that is mixed with the personal domain to avoid ambiguity. Our domain-specific findings are noteworthy given their compatibility with social domain theory despite the general nature of our decision-making items, yet replication of these longitudinal findings with other measures would increase confidence in the results.

Other studies have pointed to variation in autonomy reports depending on which family member's perspective is examined (Smetana, 1995; Smetana et al., 2004), and this is an area also in need of further investigation. Although a strength of our study is the combined use of both mother and father reports, as clearly both mothers and fathers are instrumental in negotiating youth decision-making autonomy and related family processes (cf. Parke & Buriel, 2006), the combined index of mothers and fathers may obscure important variability between parents in their perspectives on youth decision-making autonomy. Much of the previous research on decision-making autonomy has used adolescents' reports (e.g., Dornbusch et al., 1990; Fuligni & Eccles, 1993; Lamborn et al., 1996), yet further investigation of the congruence between mother, father, and adolescent reports could shed light on the family processes associated with negotiations of decision-making autonomy (see Smetana et al., 2004; Smetana et al., 2005).

Youth decision-making autonomy is likely to be best understood when studied from an ecological perspective and in cultural context. Thus, future research should implement longitudinal within-family designs to study a more diverse array of families than we were able to study here. Findings from our middle- and working-class European American sample may not generalize to other racial, ethnic, social class, regional, or cultural groups or to dynamics in single-parent or multi-generational families. A burgeoning body of research has begun to consider adolescent autonomy in a wide range of cultural contexts (e.g., Huiberts, Oosterwegel, VanderValk, Vollebergh, & Meeus, 2006; Love & Buriel, 2007; Phinney, Kim-Jo, Osorio, & Vilhjalmsdottir, 2005). Longitudinal work is now needed to examine the developmental patterns of decision-making autonomy in these various family and cultural contexts, taking advantage of designs that permit between- as well as within-family comparisons.

References

Ansbacher, H. L., & Ansbacher, R. R. (1956). *The individual psychology of Alfred Adler*. New York: Basic Books.

Arnett, J. (2000). Emerging adulthood: A theory of development from the late teens through the twenties. *American Psychologist*, *55*, 469–480.

Baumrind, D. (1991). The influence of parenting style on adolescent competence and substance use. *Journal of Early Adolescence*, *11*, 56–95.

Baumrind, D. (2005). Patterns of parental authority and adolescent autonomy. *New Directions for Child and Adolescent Development*, *108*, 61–69.

Bumpus, M. F., Crouter, A. C., & McHale, S. M. (2001). Parental autonomy granting in adolescence: Exploring gender differences in context. *Developmental Psychology*, *37*, 163–173.

Collins, W. A., Gleason, T., & Sesma, A. (1997). Internalization, autonomy, and relationships: Development during adolescence. In J. E. Grusec & L. Kuczynski (Eds.), *Parenting and children's internalization of values: A handbook of contemporary theory* (pp. 78–99). New York: Wiley.

Cox, M. J., & Paley, B. (1997). Families as systems. *Annual Review of Psychology*, *48*, 243–267.

Crouter, A. C., & Booth, A. (Eds.). (2003). *Children's influence on family dynamics: The neglected side of family relationships*. Mahwah, NJ: Erlbaum.

Crouter, A. C., Bumpus, M. F., Davis, K. D., & McHale, S. M. (2005). How do parents learn about adolescents' experiences? Implications for parental knowledge and adolescent risky behavior. *Child Development*, *76*, 869–882.

Crouter, A. C., Whiteman, S. D., McHale, S. M., & Osgood, D. W. (2007). Development of gender attitude traditionality across middle childhood and adolescence. *Child Development*, *78*, 911–926.

Daddis, C., & Smetana, J. (2005). Middle-class African American families' expectations for adolescents' behavioural autonomy. *International Journal of Behavioral Development*, *29*, 371–381.

Dornbusch, S. M., Carlsmith, J. M., Bushwall, S. J., Ritter, P. L., Leiderman, H., Hastorf, H., et al. (1985). Single parents, extended households, and the control of adolescents. *Child Development*, *56*, 326–341.

Dornbusch, S. M., Ritter, P. L., Mont-Reynaud, R., & Chen, Z. (1990). Family decision making and academic performance in a diverse high school population. *Journal of Adolescent Research*, *5*, 143–160.

Dowdy, B. B., & Kliewer, W. (1998). Dating, parent-adolescent conflict, and behavioral autonomy. *Journal of Youth and Adolescence*, *27*, 473–492.

Eccles, J. S., Buchanan, C. M., Flanagan, C. A., Fuligni, A., Midgley, C., & Yee, D. (1991). Control versus autonomy during early adolescence. *Journal of Social Issues*, *47*, 53–68.

Erikson, E. H. (1968). *Identity: Youth and crisis*. New York: W. W. Norton.

Feldman, S. S., & Wood, D. N. (1994). Parents' expectations for preadolescent sons' behavioral autonomy: A longitudinal study of correlates and outcomes. *Journal of Research on Adolescence*, *4*, 45–70.

Flanagan, C. (1990). Change in family work status: Effects of parent-adolescent decision making. *Child Development*, *61*, 163–177.

Fuligni, A. J., & Eccles, J. S. (1993). Perceived parent-child relationships and early adolescents' orientation toward peers. *Developmental Psychology*, *29*, 622–632.

Gutman, L. M., & Eccles, J. S. (2007). State-environment fit during adolescence: Trajectories of family relations and adolescent outcomes. *Developmental Psychology*, *43*, 522–537.

Hill, J. P., & Holmbeck, G. N. (1987). Disagreements about rules in families with seventh-grade girls and boys. *Journal of Youth and Adolescence*, *16*, 221–246.

Holmbeck, G. N., Paikoff, R. L., & Brooks-Gunn, J. (1995). Parenting adolescents. In M. H. Bornstein (Ed.), *Handbook of parenting: Vol. 1. Children and parenting* (pp. 91–118). Hillsdale, NJ: Erlbaum.

Huiberts, A., Oosterwegel, A., VanderValk, I., Vollebergh, W., & Meeus, W. (2006). Connectedness with parents and behavioural autonomy among Dutch and Moroccan adolescents. *Ethnic and Racial Studies*, *29*, 315–330.

Jacobs, J. E., & Klaczynski, P. A. (2002). The development of judgment and decision making during childhood and adolescence. *Current Directions in Psychological Science*, *11*, 145–149.

Kerns, K. A., Aspelmeier, J. E., Gentzler, A. L., & Grabill, C. M. (2001). Parent-child attachment and monitoring in middle childhood. *Journal of Family Psychology*, *15*, 69–81.

Kohn, M. L. (1977). *Class and conformity: A study in values* (2nd ed.). Chicago: University of Chicago Press.

Lamborn, S. D., Dornbusch, S. D., & Steinberg, L. (1996). Ethnicity and community context as moderators of the relations between family decision making adolescent adjustment. *Child Development*, *67*, 283–301.

Lareau, A. (2003). *Unequal childhoods: Class, race, and family life*. Berkeley: University of California Press.

Love, J. A., & Buriel, R. (2007). Language brokering, autonomy, parent-child bonding, biculturalism, and depression: A study of Mexican American adolescents from immigrant families. *Hispanic Journal of Behavioral Sciences*, *29*, 472–491.

Maccoby, E. E., & Martin, J. A. (1983). Socialization in the context of the family: Parent–child interaction. In P. H. Mussen (Series Ed.) & E. M. Hetherington (Vol. Ed.), *Handbook of child psychology: Vol. 4. Socialization, personality, and social development* (pp. 1–102). New York: Wiley.

McHale, S. M., Kim, J., & Whiteman, S. D. (2006). Sibling relationships in childhood and adolescence. In P. Noller & J. Feeney (Eds.), *Close relationships: Functions, forms, and processes* (pp. 127–150). Hove, UK: Psychology Press.

Nucci, L. P. (2001). *Education in the moral domain*. New York: Cambridge University Press.

Nucci, L., Camino, C., & Sapiro, C. M. (1996). Social class effects on Northeastern Brazilian children's conceptions of areas of personal choice and social regulation. *Child Development*, *67*, 1223–1242.

Parke, R. D., & Buriel, R. (2006). Socialization in the family: Ethnic and ecological perspectives. In N. Eisenberg, W. Damon, & R. Lerner (Eds.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (6th ed., pp. 429–504). New York: Wiley.

Peterson, G. W., Bush, K. R., & Supple, A. (1999). Predicting adolescent autonomy from parents: Relationship connectedness and restrictiveness. *Social Inquiry*, *69*, 431–457.

Phinney, J. S., Kim-Jo, T., Osorio, S., & Vilhjalmsdottir, P. (2005). Autonomy and relatedness in adolescent-parent disagreements: Ethnic and developmental factors. *Journal of Adolescent Research*, *20*, 8–39.

Shanahan, L., McHale, S. M., Crouter, A. C., & Osgood, D. W. (2007). Warmth with mothers and fathers from middle childhood to late adolescence: Within- and between-families comparisons. *Developmental Psychology*, *43*, 539–550.

Small, S. A., Eastman, G., & Cornelius, S. (1988). Adolescent autonomy and parental stress. *Journal of Youth and Adolescence*, *17*, 377–391.

Smetana, J. G. (1988). Concepts of self and social convention: Adolescents' and parents' reasoning about hypothetical and actual family conflicts. In M. R. Gunnar & W. A. Collins (Eds.), *Minnesota symposium on child psychology: Vol. 21. Development during the transition to adolescence* (pp. 79–122). Hillsdale, NJ: Erlbaum.

Smetana, J. G. (1995). Conflict and coordination in adolescent–parent relationships. *Child Development*, *66*, 299–316.

Smetana, J. G. (2000). Middle class African-American adolescents' and parents' conceptions of parental authority and parenting practices: A longitudinal investigation. *Child Development*, *71*, 1672–1686.

Smetana, J. G., Campione-Barr, N., & Daddis, C. (2004). Longitudinal development of family decision making:

Defining healthy behavioral autonomy for middle-class African American adolescents. *Child Development, 75,* 1418–1434.

Smetana, J., Crean, H. F., & Campione-Barr, N. (2005). Adolescents' and parents' changing conceptions of parental authority. *New Directions for Child and Adolescent Development, 108,* 31–46.

Steinberg, L. (2001). We know some things: Parent–adolescent relationships in retrospect and prospect. *Journal of Research on Adolescence, 11,* 1–19.

Steinberg, L., Elmen, J. D., & Mounts, N. S. (1989). Authoritative parenting, psychosocial maturity, and academic success among adolescents. *Child Development, 60,* 1424–1436.

Sulloway, F. (1996). *Born to rebel: Birth order, family dynamics, and creative lives.* New York: Vintage Books.

Turiel, E. (1998). The development of morality. In W. Damon (Series Ed.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (5th ed., pp. 863–962). New York: Wiley.

Whiteman, S. D., McHale, S. M., & Crouter, A. C. (2003). What parents learn from experience: The first child as a first draft? *Journal of Marriage and Family, 65,* 608–621.

Whiteman, S. D., McHale, S. M., & Crouter, A. C. (2007). Longitudinal changes in marital relationships: The role of offspring's pubertal development. *Journal of Marriage and Family, 69,* 1005–1020.

Child Development, March/April 2010, Volume 81, Number 2, Pages 652–668

Children's Sensitivity to the Conventionality of Sources

Gil Diesendruck and Nurit Carmel
Bar-Ilan University

Lori Markson
Washington University

Four studies examined preschoolers' sensitivity to agents' knowledge of conventional forms. Three- to 4-year-olds heard a speaker apply either conventional or wrong labels to familiar objects (Studies 1 and 2, *N* = 57) or peculiar but correct labels (Study 3, *N* = 19). When then asked by the speaker for the referent of a novel label, children exposed to an accurate labeler were more likely to choose an unfamiliar object than children exposed to an inaccurate labeler. Study 4 (*N* = 36) replicated these findings using object functions instead of labels. Children hold an assumption of conventionality with regard to both object labels and functions, but they are selective in their application of this assumption toward agents who are knowledgeable of conventions in these domains.

An implicit assumption adults make when communicating is that their partners in a conversation share with them the knowledge that there are certain preferred linguistic forms to express certain meanings. In fact, adults expect their partners to use these forms when intending to express certain meanings and consequently draw inferences about their partners' communicative intents based on their choice of forms. Clark (1988, 1990) defined this basic assumption as the principle of conventionality, and an ensuing inference based on the use of nonconventional forms as the principle of contrast—namely, that if a speaker uses a linguistic form that is different from the conventional one expected, it is probably because the speaker has a contrastive intent in mind.

A number of recent studies show that from a young age, children too seem to rely on an assumption of conventionality when interpreting people's utterances. In fact, studies indicate that already by 3 years of age, children manifest a nuanced assumption of conventionality, such that they are capable of modulating when and when not to apply this assumption (see Kalish & Sabbagh, 2007, for a review).

Most studies of children's assumption of conventionality regarding words have investigated it in the context of its contrastive implication. A widely studied situation in which this implication is manifest is when children are presented with two objects, one familiar and one unfamiliar, and are simply asked to point to the referent of an unfamiliar noun. Based on a principle of conventionality, children in such a situation should assume that the speaker knows the conventional name of the familiar object, and they should thus expect speakers to use that name if he or she intends to refer to that object. According to a principle of contrast, children should reason that if a speaker uses a different unfamiliar name, it is probably because she has a different object in mind. That is, children should select the unfamiliar object as the referent of the unfamiliar name. And indeed, that is what a number of studies with preschool children have found (Markman & Wachtel, 1988; Merriman & Bowman, 1989).

In itself, however, this type of response is also consistent with proposals arguing that children have a number of constraints that guide their word learning. In particular, the response is consistent with a mutual exclusivity bias (Markman & Wachtel, 1988) and a novel-name nameless-category principle (Golinkoff, Mervis, & Hirsh-Pasek, 1994). Two unique predictions of the pragmatic reasoning process exposed above are that children's inferences of contrast should be contingent on:

This research was partly funded by an internal grant from Bar-Ilan University's Rector's Office. We would like to thank the teachers, parents, and children for their participation in the studies. We are also grateful to the following research assistants for their help with data collection and coding: Itai Cohen, Anat Grosz, Elisheva Hecht, Avital Rozensweig, Yael Tadmor, and Amiad Tauber.

Correspondence concerning this article should be addressed to Gil Diesendruck, Department of Psychology, Bar-Ilan University, Ramat-Gan, 52900 Israel. Electronic mail may be sent to dieseng@mail.biu.ac.il.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8102-0018

(a) whether or not the *form* used is thought to be conventional and (b) whether or not the *speaker* or *user* is believed to know the form.

Regarding the forms about which children assume conventionality, studies have found that children believe that common nouns are known by other speakers of their language, but arbitrary facts about objects, proper nouns, or people's preferences and goals are not. For instance, in Diesendruck and Markson (2001), an experimenter showed 3- to 4-year-old children two novel objects and taught them either a novel label or a novel fact for one of the objects. A puppet, who was either present or absent during this introduction, then asked children to select the object associated with a different novel label or fact, accordingly. Diesendruck and Markson found that children avoided assigning two facts for an object only when the puppet was present during the experimenter's introduction of the fact but not when he was absent. In contrast, children avoided assigning two labels for an object both when the puppet was present and when he was absent. In fact, when directly asked whether or not the puppet knew the labels given by the experimenter, most children replied that the absent puppet did. These findings suggest that children presuppose that speakers of a language know the names of objects in their language but hold no such presupposition about arbitrary facts associated with objects. Children assumed the puppet knew the conventional names of objects and inferred that if he asked them for the referent of a different name it was because he had a different object in mind. Using a similar methodology, Diesendruck (2005) found that preschoolers assume that speakers know the common nouns associated with objects but not the proper nouns (see also Birch & Bloom, 2002). In other words, only certain words are believed to be conventional, thus warranting an inference of contrastive intents. Finally, Henderson and Graham (2005) found that 2-year-olds assume absent adults share with an experimenter knowledge of the novel names of objects but not the experimenter's preferences toward the objects (see also Buresh & Woodward, 2007; Graham, Stock, & Henderson, 2006, for similar findings with younger children).

Regarding the second pragmatic prediction mentioned above, there are a number of studies indicating that already by 2 years of age, children are sensitive to speakers' knowledge state when interpreting their utterances (Akhtar, Carpenter, & Tomasello, 1996; Diesendruck, Markson, Akhtar, & Reudor, 2004; Sabbagh & Baldwin, 2001). Only one study, however, has attempted to address whether children's inference of contrast is modulated by speakers' presumed knowledge of language. Diesendruck (2005) presented pairs of novel objects to Hebrew-English bilingual 3- to 4-year-olds. An experimenter labeled one of the objects with a novel English label, and then a puppet—who was present during the labeling event—asked children for the referent of a novel Hebrew label. The crucial manipulation was that for half of the children, the puppet was introduced—and behaved—as a monolingual Hebrew speaker, and for the other half the puppet was introduced as being a Hebrew-English bilingual. Diesendruck found that children selected the yet unlabeled object when the puppet was bilingual but selected randomly between the two objects when the puppet was monolingual. In other words, children seem to have reasoned that if a monolingual speaker of Hebrew hears an experimenter teach them an English novel word, then the monolingual speaker does not necessarily know that foreign word. Consequently, if that speaker asks children for the referent of a novel Hebrew word, they infer that the speaker might be referring to either object—not necessarily the object that did not have an English label applied to it. In turn, if a bilingual speaker who has heard the experimenter use an English label chooses to use a different label—even if in a different language—then children assume that he knows that English label and thus take the speaker's use of a different label to imply a contrastive intent.

Whereas the findings from this latter study are consistent with the prediction that children assume conventionality only with respect to people who are knowledgeable about the conventions of a language, children in the above study did not get any *direct* evidence for the speakers' knowledge of the relevant conventions. In fact, the above study did not manipulate whether or not the speakers were knowledgeable about *conventional* linguistic forms but rather whether or not the speakers were knowledgeable about the *language* in general. In other words, we cannot determine unambiguously from that study that children's responses derived specifically from the fact that the monolingual speaker violated conventionality per se. The rationale behind Diesendruck's (2005) study was that children might have presumed that a monolingual Hebrew speaker would not have understood the English interaction between the experimenter and the child, and thus would not have learned the *novel label* the experimenter taught the child to one of the objects. The study did not assess, however, whether children believed either speaker knew the *familiar labels* of objects. A primary goal of the present studies was to provide children with specific evidence that certain speakers of their language do not know the conventional names of objects and that, consequently, children should withhold from applying an assumption of conventionality toward such speakers. Thus, different from the Diesendruck (2005) study, the present studies attempted to manipulate only the most relevant aspect of knowledge presupposed by the conventionality account of the disambiguation response.

The second major goal of the studies was to extend the examination of the scope of an assumption of conventionality. As reviewed above, originally, studies on children's assumption of conventionality targeted primarily the domain of language. In fact, there are some suggestions that such an assumption derives from the manner in which parents talk to children (Callanan & Sabbagh, 2004). Like linguistic forms, however, certain cultural forms are also conventionalized. For instance, the functions we attribute to objects (e.g., what forks are for) and the manner in which we use objects (e.g., how we hold forks) are different types of culture-specific knowledge that members of a community need to know and expect peers to know. In recent years, a number of researchers have suggested that, in addition to the domain of language, an assumption of conventionality seems to hold in domains such as games (Rakoczy, 2007) and object functions (German, Truxaw, & Defeyter, 2007; Siegel & Callanan, 2007). For instance, Casler and Kelemen (2005, 2007) showed that 2½-year-olds assume that objects have unique functions. In particular, upon seeing an object being used intentionally for a specific function, even 2-year-olds readily picked that object over other equally affording objects to perform that function. These studies, however, have been conducted with a variety of methodologies and age groups (see Kalish & Sabbagh, 2007, for a review) and thus do not allow for direct comparisons between domains. The present studies used almost identical methods to evaluate whether children hold an assumption of conventionality about both object labels and object functions, and whether they draw similar inferences—namely, about contrast—based on people's manifest knowledge of conventional labels and functions.

In order to address both goals described above, we adopted a methodology originally developed by Koenig and Harris (2005; Koenig, Clement, & Harris, 2004; see also, Birch, Vauthier, & Bloom, 2008; Jaswal & Neely, 2006). In Koenig et al. (2004), for instance, children were either exposed to speakers who manifested knowledge of the conventional names of objects—so-called reliable speakers—or to speakers who manifested ignorance of conventional names—so-called unreliable speakers. Koenig et al. found that 3- to 4-year-olds were more likely to subsequently learn a novel name from a knowledgeable speaker than from an ignorant one (see also Sabbagh & Baldwin, 2001). Different from all these previous studies, in the present studies, rather than assessing whether the knowledge states of speakers differentially influence children's *willingness to learn* labels from such speakers, we asked whether children within this same age range would *draw different inferences* about such speakers' communicative intents in an ambiguous context. That is, we asked whether variations in speakers' manifest knowledge of conventional forms influence the extent to which 3- to 4-year-olds respond to speakers in the manner stipulated by an assumption of conventionality.

To address the first goal, in Study 1 we familiarized 3- to 4-year-old children with two different kinds of speakers: a speaker who showed evidence of being knowledgeable about the conventional labels used to refer to familiar objects and a speaker who showed evidence of being ignorant of the conventional labels used to refer to familiar objects. Children then participated in a standard disambiguation task, in which they were presented with pairs of objects consisting of one familiar and one unfamiliar object, and the same speaker to which they had been previously exposed asked them for the referent of a novel label. The rationale is that if children's inference of contrast is contingent on them assuming that the speaker knows the conventional forms of the language, then children who had been exposed to a knowledgeable speaker should be more likely than those exposed to an ignorant speaker to select the unfamiliar object as the referent of the novel label. Study 2 replicates Study 1 with a different operationalization of speakers' knowledge. Study 3 examines a further implication of this hypothesized reasoning process.

In order to address the second goal, in Study 4 the same method described above was employed, but instead of exposing children to speakers and object labels, we exposed children to users and object functions. A recent study using a similar methodology found that just as with labels, children are more willing to learn a new function from a knowledgeable labeler of object functions than from an ignorant one (Birch et al., 2008). Here, we ask whether users' knowledge affects children's inferences about the conventionality of users. As in the first three studies, we first familiarized children with two kinds of users: a user who showed evidence of being knowledgeable about the conventional functions of familiar objects and a user who showed evidence of being ignorant about the conventional functions of familiar objects. Children then participated in a function disambiguation task, in which they were presented with pairs of objects consisting of one familiar and one unfamiliar object, and the same user to which they had been previously exposed asked them for the object to be used for a new function. The rationale is that if children infer the functions of objects via the same reasoning process deployed for object labels, then in the function disambiguation task, they should show the same differential responding to a knowledgeable and an ignorant user as that hypothesized for knowledgeable and ignorant speakers.

Study 1

Method

Participants

Participants were 33 monolingual Hebrew-speaking Israeli preschoolers, 19 boys and 14 girls (M = 3 years 10 months, SD = 4 months, range = 3 years 3 months to 4 years 5 months). Children were primarily from middle- and upper-middle-class socioeconomic backgrounds. Most children were recruited from and tested in preschools around the university. A few others were recruited from the local community and were tested at their homes. Parental consent was obtained for each child prior to participation.

Materials

In the familiarization phase, four objects highly familiar to children this age were used. The objects were: a plastic banana, a stuffed dog, a doll-size shirt, and a plastic cookie. In the disambiguation phase, eight objects were used: four familiar and four novel. The objects were arranged into pairs of familiar–novel, in such a way that the objects in each pair were similar in their color salience and size. The familiar–novel pairs were: (a) shoe–funnel, (b) car–door-stopper, (c) cup–cylinder, and (d) pen–original object created in the lab. A hand-puppet was also used to provide the labels. The experimenter manipulated the puppet, using a distinctive adult voice whenever speaking for him.

Design

Children were randomly assigned to one of two conditions, *knowledgeable* (n = 16, 7 girls and 9 boys, M_{age} = 3 years 10 months) or *ignorant* (n = 17, 7 girls and 10 boys, M_{age} = 3 years 9 months). There was neither a significant age difference nor a significant difference in the gender distribution between the two conditions. Children participated in four familiarization trials followed by four disambiguation trials.

Procedure

Prior to testing children, the experimenter came to the preschool and played with children for at least 1 hr, so that children would become comfortable with the experimenter. Once rapport had been established, the experimenter invited each child to come sit with her in a quiet area set up ahead of time in the preschool. The entire interaction—in this and all the other studies—was conducted in Hebrew (verbatim translations of what children were told are provided). The child and the experimenter sat at a table, and the experimenter introduced children to her "friend"—a hand-puppet named "Yossi." The experimenter exchanged a few greeting words with the puppet, speaking with a lower toned voice when speaking for the puppet. This introduction was meant to establish that the puppet could speak and was separate from the experimenter. After this brief introduction—which was identical for children in both conditions—the experimenter moved on to the familiarization phase.

Familiarization phase. The familiarization phase was the only phase that differed between the two conditions. In this phase, the experimenter invited the child and the puppet to play with some objects she had brought. The experimenter pulled out one object at a time from her bag and laid it on the table. The puppet looked at the object and named it three times while asserting his knowledge of what the object is (i.e., saying, "I know what this is!"). The main manipulation had to do with whether or not the labels provided by the puppet were correct. In the knowledgeable condition, the puppet labeled the objects correctly with their conventional basic-level names (e.g., upon seeing a banana, he said, "That's a banana, it's a banana; look at the banana!"). In the ignorant condition, the puppet labeled the objects incorrectly (e.g., called the banana an "apple"). After a few seconds during which the child and the puppet could play with an object, the experimenter returned the object to her bag and brought out another familiar object. Again the puppet reacted according to the child's condition. This procedure was repeated for four trials, in order to allow the child to build an assessment of the puppet's knowledge—or ignorance—of the familiar objects' names. The correct labels used for the items banana, dog, shirt, and cookie, were *banana*, *dog*, *shirt*, and *cookie*, respectively. The incorrect labels used were: *apple*, *cow*, *pants*, and *egg*, respectively. Objects were presented in four different orders, such that each object appeared in each position the same number of times across children.

Disambiguation phase. This phase was identical in both conditions. The experimenter placed the first pair of familiar–novel objects in front of the child. The puppet then looked at the child, and asked for one of the objects by using a Hebrew-sounding made-up novel label (e.g., "Can you give me the *beega*?" the other three novel labels were: *bakakhat*, *shafek*, and *geet*). Both the puppet and the experimenter refrained from looking at either one of the objects and instead maintained their gaze on the child. Once the child responded by pointing to, picking up, or handing to the puppet one of the objects, the experimenter removed the two objects, and replaced them with the next pair of familiar–novel objects. The experimenter did not provide any feedback to children about their performance, and instead commented on the game and the toys (e.g., "It's fun to play with these things, right?"). The experimenter recorded children's choices and any other verbal comments children made throughout the entire procedure. Upon completion of the four disambiguation trials, children were thanked, given a sticker in appreciation for their participation, and returned to their class.

The order of presentation of the object pairs was counterbalanced across children, such that each object pair appeared an equal number of times as the first, second, third, and fourth. The left–right placement of each object within a pair was counterbalanced within children, so that on half of the pairs, the familiar object was placed on the right side of the child, and on the other half of the pairs, the novel object was placed on the right side of the child.

Results

The main prediction was that children would be more likely to choose the novel object in response to a request made by a knowledgeable speaker than to one made by an ignorant speaker. The dependent measure used to address this prediction was the number of times (0–4) children selected the novel object as the referent of the novel label used by the puppet in the disambiguation phase.

An analysis of variance (ANOVA) with condition (knowledgeable vs. ignorant) and gender as between-subjects variables, and the participants' age in months as a covariate, on the above dependent measure revealed a main effect of condition, $F(1, 28) = 5.17$, $p < .05$, $\eta_p^2 = .16$; no effect of gender, $p > .8$; no effect of age, $p > .5$; and no interaction between gender and condition, $p > .7$. As illustrated in Figure 1, children in the knowledgeable condition ($M = 3.75$, $SD = 0.45$) were more likely than children in the ignorant condition ($M = 3.00$, $SD = 1.17$) to pick the novel object in response to the puppet's request for a novel label. Interestingly, analyses against chance (chance = 2 selections of a novel object) revealed that children in both conditions were more likely to pick the novel object than would be expected by chance: in the knowledgeable condition, $t(15) = 15.65$, $p < .001$, and in the ignorant condition, $t(16) = 3.52$, $p < .005$.

In order to evaluate children's individual pattern of responses, we classified children into two groups based on how consistently they selected the novel object. Children who selected the novel object on three or four of the four trials were considered *disambiguators*, and those who selected the novel object on two or fewer of the four trials were considered *overlap accepters*. As shown in Table 1, all of the children in the knowledgeable condition were disambiguators, whereas only 10 of the 17 children in the ignorant condition were so, $\chi^2(1, N = 33) = 8.36$, $p < .005$.

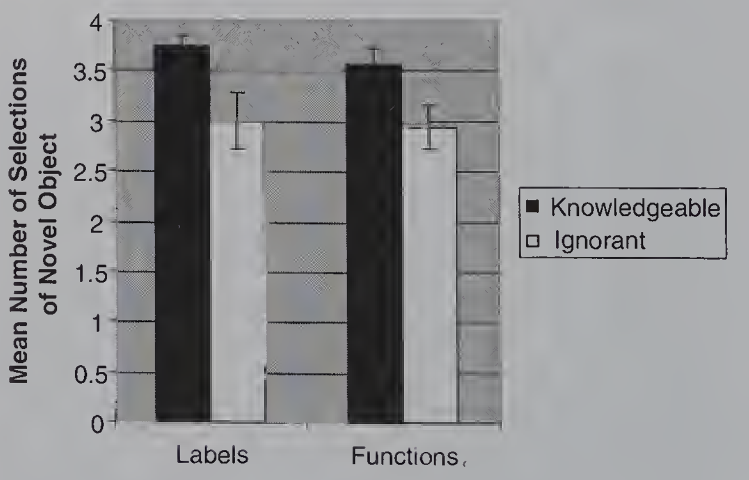


Figure 1. Mean number of selections of novel object in response to novel labels (Study 1) or new functions (Study 4).
Note. Error bars represent standard errors.

Table 1
Distribution of Children in Terms of Response Patterns Across Conditions, in Studies 1, 2, and 4

	Study 1		Study 2		Study 4	
Pattern/condition	Know	Ignor	Know	Ignor	Know	Ignor
Disambiguators	16	10	8	3	16	12
Overlap accepters	0	7	4	9	2	6
N	16	17	12	12	18	18

Note. Know = knowledgeable; Ignor = ignorant.

Finally, in order to directly address a potential interaction between age and condition, we broke down the participants into two age groups according to the median age. This new "age group" variable was then entered into an ANOVA on children's selection of the novel object, with condition as the other between-subjects variable. The interaction between these variables was not significant, $p > .7$.

Discussion

The goal of Study 1 was to investigate whether children would rely on a speaker's preestablished knowledge or ignorance of conventional object names when later interpreting his request for the referent of a novel label. We hypothesized that children exposed to a speaker who had manifested knowledge of conventional object names, compared to those exposed to a speaker who had manifested ignorance of conventional object names, would be more likely to uphold conventionality regarding such a speaker and thus expect him to use conventional labels to refer to familiar objects. Such a knowledgeable speaker's use of a nonconventional (novel) label would thus be more likely to be inferred as manifesting a contrastive referential intent than when an ignorant speaker does so. Namely, a knowledgeable speaker's, but not an ignorant speaker's, use of a novel label would likely indicate his intent to refer to an unfamiliar object.

The results of Study 1 were consistent with this hypothesis. Children were more likely to select a novel object as the referent of a novel label when the speaker asking for the novel label had previously manifested knowledge of the conventional names of objects than when the speaker manifested ignorance of conventional names. These results show that by 3 years of age, children monitor speakers' knowledge of the conventional forms existent in their language and use this information to draw inferences about speakers' referential intents. This adds to Diesendruck's (2005) findings on children's responses to bilingual versus monolingual speakers and reveals that not only can children monitor whether or not a speaker knows the language in general but they in fact react differently to a speaker of the language depending on how knowledgeable or ignorant of conventional object labels the speaker has proven to be.

A possible alternative interpretation of these findings is that children's responses did not derive from assessments of the speakers' knowledge. Rather, children might have simply noticed that one speaker uses the labels that are expected, and the other does not. Consequently, when in the disambiguation phase the ignorant speaker used an unusual label, children might have simply inferred that it is plausible that here too he was referring to the familiar object. According to this interpretation, children might have conceived of the ignorant speaker not so much as lacking knowledge of familiar object names but rather as being silly or joking. Such an interpretation might have been especially exacerbated if children thought it unlikely that a "mature" speaker of a language—as the puppet in Study 1 was portrayed to be—would not know the names of familiar objects.

In order to address this alternative interpretation, in Study 2 we attempted to create scenarios in which a speaker's ignorance would: (a) seem more ecologically plausible and, consequently, (b) less likely to be interpreted as manifesting playfulness.

Study 2

Study 2 followed nearly the identical design and procedure as Study 1 except for two modifications. First, instead of introducing the puppet as a friend of the experimenter, the experimenter described the puppets as children. Crucially, in the ignorant condition, the puppet was described as a young child who had just started learning Hebrew. In the knowledgeable condition, the puppet was described as a school-aged child who already knows a lot of Hebrew. It was expected that this modification would make the behavior of the speaker in the ignorant condition more plausible to the participants. Second, in order to directly assess children's appraisal of the speakers' behaviors, at the end of the experimental trials, the experimenter asked children explicitly about whether they thought the speaker knew the names of the familiar objects in the disambiguation phase and whether they thought the speaker was joking when he labeled the familiar objects in the familiarization phase. These questions served as manipulation checks to verify that children's answers in the two different conditions derived primarily from their assessments of speakers' knowledge states rather than differences in speakers' playfulness.

Method

Participants

Participants were 24 monolingual Hebrew-speaking Israeli preschoolers, 13 boys and 11 girls (M = 3 years 10 months, SD = 5.5 months, range = 3 years 0 months to 4 years 7 months). Children were primarily from middle- and upper-middle-class socioeconomic backgrounds. Children were recruited from, and tested in, their preschools. Parental consent was obtained for each child prior to participation.

Materials

The same materials used in Study 1 were used in Study 2.

Design

Children were randomly assigned to one of two conditions: *knowledgeable–older child* (n = 12, 6 girls and 6 boys, M_{age} = 3 years 9 months) or *ignorant–younger child* (n = 12, 5 girls and 7 boys, M_{age} = 3 years 11 months). There was neither a significant age difference nor a significant difference in the gender distribution between the two conditions. As in Study 1, children participated in four familiarization trials followed by four disambiguation trials.

Procedure

The procedure in Study 2 was almost identical to that of Study 1 except for two modifications: one having to do with the presentation of the puppets and the second having to do with the addition of two questions at the end of the procedure.

Children in the knowledgeable condition in Study 2 were introduced to a puppet described as being an older child. Children were prompted about their acquaintance with children who already attend school, and thus know a lot of Hebrew, and were told that Yossi, the puppet, is like these children. In contrast, children in the ignorant condition were told that the puppet is a young child and that he has just started learning Hebrew. After the introduction of the puppet, the procedure continued exactly as in Study 1. That is, the puppet named, either correctly (knowledgeable condition) or incorrectly (ignorant condition) the four familiar objects in the *familiarization phase* and then asked children for the referents of novel labels in the four trials of the *disambiguation phase*. The same counterbalancing precautions taken in Study 1 were undertaken in Study 2.

After children completed the disambiguation phase, the experimenter pointed to the familiar object in the last trial of that phase (e.g., a pen) and asked children whether they thought the puppet knew the name of the object (e.g., "Do you think that Yossi knows that this is called a 'pen'?"). Children's yes–no answers were recorded. The experimenter then brought out one of the familiar objects from the familiarization phase (e.g., the dog), reminded children what the puppet had called it (e.g., "cow" in the ignorant condition), and asked children if they thought the puppet was joking when he said that. Again, children's yes–no responses were recorded.

Results

The main prediction was that children would be more likely to choose the novel object in response to a request made by a knowledgeable speaker than to one made by an ignorant speaker. The dependent measure used to address this prediction was the number of times (0–4) children selected the novel object as the referent of the novel label used by the puppet in the disambiguation phase.

An ANOVA with condition (knowledgeable vs. ignorant) and gender as between-subjects variables, and the participants' age in months as a covariate, on the above dependent measure revealed a main effect of condition, $F(1, 19) = 14.90$, $p < .005$, $\eta_p^2 = .44$; an effect of gender, $F(1, 19) = 6.97$, $p < .05$, $\eta_p^2 = .27$; but no effect of age, $p > .3$, and no interaction between gender and condition, $p > .2$. The effect of gender resulted from boys ($M = 2.69$, $SD = 1.32$) being more likely than girls ($M = 1.64$, $SD = 1.36$) to choose the novel object. In terms of the condition effect, children in the knowledgeable condition ($M = 3.00$, $SD = 1.05$) were more likely than children in the ignorant condition ($M = 1.42$, $SD = 1.31$) to pick the novel object in response to the puppet's request for a novel label. Analyses against chance (chance = 2 selections of a novel object) revealed that children in the knowledgeable condition, $t(11) = 3.32$, $p < .01$, but not in the ignorant condition, $p > .1$, selected novel objects more often than expected by chance. Finally, as in Study 1, we looked at a possible interaction between age and condition by breaking down the participants into two age groups according to the median age. Here, too, we found no interaction between these factors, $p > .1$.

Analogous to Study 1, in order to evaluate children's individual pattern of responses, we classified children into two groups based on how consistently they selected the novel object. Children who selected the novel object on three or four of the four trials were considered *disambiguators*, while those who selected the novel object on two or fewer of the four trials were considered *overlap accepters*. As shown in Table 1, there were more disambiguators in the knowledgeable than in the ignorant condition, and the opposite was the case in terms of overlap accepters, $\chi^2(1, N = 24) = 4.20$, $p < .05$.

Our last set of analyses looked at children's responses to the two questions at the end of the procedure. We found that all children in the knowledgeable condition responded affirmatively to the question whether or not the puppet knew the name of the last familiar object; however, all but 1 child in the Ignorant condition, responded negatively, $\chi^2(1, N = 24) = 20.31$, $p < .001$. In turn, all children in both conditions, except for 1 in the ignorant condition, responded that the puppet was not joking when he named the familiar object in the familiarization phase.

Discussion

The results of Study 2 are consistent with our interpretation that children's responses to the requests of speakers' who had correctly versus wrongly labeled familiar objects, indeed derived from their assessments of these speakers' knowledge of conventional object names, rather than estimations of their seriousness. This finding is consistent with Koenig and Harris (2005), who found that children infrequently explained an unreliable speaker's incorrect labeling of familiar objects in terms of "pretense" or "silliness." This interpretation is also consistent with previous findings showing that children are more willing to learn new information from sources who had previously been accurate than from those who had been inaccurate (e.g., Birch et al., 2008; Koenig & Harris, 2005). Moreover, the finding that children were capable of differentiating explicitly between a knowledgeable and an ignorant puppet indicates that their responses indeed reflected their appraisal of the puppets, not of the experimenter manipulating the puppets.

An unexpected finding from Study 2 is that boys were more likely than girls to select the novel object. We do not have a clear account of this gender effect. In fact, given that this effect occurred only in this study, it might have been due to some peculiarity in the sample. Moreover, the fact that gender did not interact with condition, that is, both boys and girls were more likely to select the novel object in response to a request made by a knowledgeable than by an ignorant speaker, suggests that any difference there might exist between the genders, it does not seem to involve their capacity to assess speakers' knowledge.

One difference between the findings in Study 2 and those in Study 1 is that whereas children in the ignorant condition of Study 2 selected the novel objects at rates no different from chance, children in the ignorant condition of Study 1 selected the novel objects at rates significantly above chance. Interestingly in this regard, children in Diesendruck's (2005) study, who were exposed to a monolingual speaker who did not know anything about the language used previously to label an object, also responded at chance levels. A possible reconciliation of these findings has to do with the inferences children might have made about the different kinds of speakers. Namely, it is possible that an encounter with a mature speaker who knows the language but nonetheless commits errors in naming familiar objects (i.e., the ignorant speaker in Study 1), was quite puzzling to children. Such a speaker may have prompted children to doubt his knowledge of other familiar words, and the subsequent reasoning process children needed to engage in so as to fix this skepticism may have led to decrement in disambiguation responses. In this light, the encounter with an *immature* speaker who committed errors in naming familiar objects (i.e., the ignorant speaker in Study 2) might have been less puzzling and thus easier to process. Specifically, children might have felt more comfortable simply attributing ignorance to such a speaker, thus not disambiguating at above-chance levels.

A second possible reconciliation has to do with the identity of the speakers. Namely, children may assume that young children or speakers of a different language indeed know very little, if anything, about their language, but presumably mature speakers, even if they have been temporarily inaccurate, may nonetheless know more about the language. In other words, children may be relatively tolerant in their upholding of an assumption of conventionality, and it is only when a speaker is immature or does not understand the language at all that they more readily abandon this assumption (see also Lampinen & Smith, 1995, for consistent findings).

One implication of this latter interpretation is that if a mature speaker manifests knowledge of the language—even if the manner in which he or she refers to objects is unusual—then children should nonetheless uphold an assumption of conventionality about him or her. Study 3 addresses this implication.

Study 3

In order to assess how tolerant children's assumption of conventionality is, in Study 3 we exposed children to a speaker (a puppet friend of the experimenter, as in Study 1) who employed unusual but correct labels to refer to familiar objects. In other words, this type of speaker gave evidence that he was knowledgeable about the language in general, even if his choice of labels to refer to the objects in the familiarization phase was peculiar. To ascertain that children were familiar with these unusual labels and could assess their accuracy, the superordinate labels for each of the familiar objects was used as the unusual label. While familiar to children of this age, superordinate labels are much less frequent in the linguistic input to which children are typically exposed, and are used less frequently by children (Callanan, 1985; Hall, 1994).

The alternative predictions were that if children hold a rigid assumption of conventionality (i.e., that each object has *only one* correct label—typically its basic-level label), then they should infer that an unusual labeler does not know the names of familiar objects, and they should thus respond like children in the Ignorant condition of Study 1 did. In turn, if children hold a more tolerant assumption of conventionality—as hinted by the fact that even children in the ignorant condition of Study 1 selected the novel object at above chance levels—then they should infer that speakers who are knowledgeable about the language might know the conventional labels of familiar objects. In this case, children should respond to Study 3's unusual but knowledgeable labeler like children in the knowledgeable condition of Study 1 did.

Method

Participants

Participants were 19 monolingual Hebrew-speaking Israeli preschoolers, 12 boys and 7 girls (M = 3 years 10 months, SD = 4.7 months, range = 3 years 4 months to 4 years 9 months). Children were primarily from middle- and upper-middle-class socioeconomic backgrounds. Most children were recruited from and tested in preschools around the university. A few others were recruited from the local community and were tested at home. Parental consent was obtained for each child prior to participation.

Materials

The same materials used in Studies 1 and 2 were used in Study 3.

Design

All 19 children participated in the *unusual* condition, in which they were introduced to a puppet in the same way as they had been in Study 1 (i.e., the puppet was introduced as a friend of the experimenter, not as a child). As in Studies 1 and 2, the procedure consisted of four familiarization trials followed by four disambiguation trials.

Procedure

The procedure in Study 3 was nearly identical to the one used in Study 1 with only one exception. In the *familiarization phase,* children were exposed to a puppet who consistently used unusual but correct labels to name the four familiar objects (the same objects used in Study 1). In particular, the puppet used superordinate labels and thus called the banana *fruit*, the dog *animal*, the shirt *clothes* (in Hebrew, there is a singular form for clothes), and the cookie *food*. The rest of the procedure, including the *disambiguation phase* and the counterbalancing precautions, was identical to the procedure in Study 1.

Results

As in Study 1, the dependent measure used to assess children's responses was the number of times (0–4) children selected the novel object as the referent of the novel label used by the puppet in the disambiguation phase. In general, children in Study 3 selected the novel object an average of 3.79

times ($SD = 0.42$), which was significantly more than would be expected by chance (chance = 2), $t(18) = 18.62$, $p < .001$.

In order to address the alternative predictions described above, two planned comparisons were conducted. First, we compared children's performance in Study 3 with the performance of children in Study 1's knowledgeable condition. This comparison revealed no significant difference, $p > .7$. Next, we compared children's performance in Study 3 with the performance of children in Study 1's ignorant condition. This comparison revealed that indeed, children in the unusual condition of Study 3 were more likely to select the unfamiliar object in response to the novel label than were children in the ignorant condition of Study 1, $t(34) = 2.75$, $p < .01$.

As in Study 1, we also conducted a nonparametric test in order to evaluate differences in individual children's response patterns. Using the same classificatory criterion described in Studies 1 and 2, we divided children in the unusual condition into *disambiguators* and *overlap accepters*. All 19 children in the unusual condition were disambiguators. This distribution was similar to the one found among children in the knowledgeable condition of Study 1 and was significantly different from the distribution found in the ignorant condition of Study 1, $\chi^2(1, N = 36) = 9.71$, $p < .005$.

Taken together, the findings from Study 3 reveal that children responded to an unusual but knowledgeable speaker in the same way that they responded to a knowledgeable speaker who used typical basic-level labels to name familiar objects.

Discussion

The goal of Study 3 was to examine whether children would uphold an assumption of conventionality toward a speaker who—while manifesting knowledge of the language—had used unusual names to refer to familiar objects. The results of Study 3 indicate that children did just that. Children seemed to have assumed that if a speaker knows that a banana is called *fruit*, then he likely also knows what a car and a shoe are called. Consequently, when he asked children for the referent of a completely novel word, children inferred that he must be referring to a novel object rather than to one for which children knew the name. This finding also undermines an associationist account of the findings in Study 1: Children did not think that an unusual labeler would always refer to familiar objects with unusual names.

These findings also point to a further implication of conventionality, namely, that while there are *typical* ways of referring to objects, there is not necessarily *only one correct way*. As Clark (1997) discusses, adults and children are aware of the multiple perspectives one can take to refer to objects. We will return to this point in the General Discussion.

Study 4

The previous three studies revealed that children rely on a speaker's manifested knowledge or ignorance of the conventional names of objects when subsequently inferring the speaker's referential intent. In other words, when children have sufficient evidence to uphold an assumption of conventionality toward a speaker, they feel licensed to draw contrastive inferences. This conclusion adds to a growing number of studies revealing children's reliance on an assumption of conventionality regarding words (Buresh & Woodward, 2007; Callanan & Sabbagh, 2004; Diesendruck & Markson, 2001; Henderson & Graham, 2005).

As noted in the Introduction, an assumption of conventionality is useful not only in the realm of language but also with regard to other cultural forms that although somewhat arbitrary and symbolic, are also conventionalized within cultural groups (Kalish & Sabbagh, 2007). The goal of Study 4 was to examine the scope of an assumption of conventionality by investigating whether young children hold such an assumption with regard to object functions. There is evidence that children believe that objects have unique functions (Casler & Kelemen, 2005, 2007) and that these functions are the ones shared by multiple people (Siegel & Callanan, 2007). A recent study also demonstrated that just as with labels, 3- and 4-year-olds prefer learning about the function of a novel object from a user who had been previously accurate in describing the uses of objects than from a user who had been faulty (Birch et al., 2008). In Study 4, we wanted to directly compare children's assumptions and subsequent inferences about names and functions. Study 4 adopts the same methods used in Study 1 to assess children's reasoning process in the domain of names in order to assess their reasoning in the domain of object functions.

Method

Participants

Participants were 36 monolingual Hebrew-speaking Israeli preschoolers, 14 boys and 22 girls (*M* = 3 years 11 months, *SD* = 6.2 months, range = 3 years 3 months to 4 years 8 months). Children were primarily from middle- and upper-middle-class socioeconomic backgrounds. Most children were recruited from and tested in preschools around the university. A few others were recruited from the local community and were tested at home. Parental consent was obtained for each child prior to participation.

Materials

In the familiarization phase, we used four objects highly familiar to children this age and whose function children were assumed to know. The objects were: a toy chair, a plastic spoon, a wooden hammer, and a plastic comb. In the disambiguation phase, eight objects were used: four familiar and four novel. The objects in the disambiguation phase were arranged into pairs of familiar–novel, in such a way that the objects in each pair were similar in their color salience and size, and both objects could fulfill the new functions mentioned in this phase. The familiar–novel pairs are described in Table 2. The pairs were the same for all children but were presented in different orders across children. A hand-puppet was also used to provide the functions. The experimenter manipulated the puppet, using a distinctive voice whenever speaking for him (as in Study 1).

Design

Children were randomly assigned to one of two conditions, *knowledgeable* ($n = 18$, 12 girls and 6 boys, M_{age} = 3 years 11 months) or *ignorant* ($n = 18$, 10 girls and 8 boys, M_{age} = 4 years 0 months). There was neither a significant age difference nor a significant difference in the gender distribution between the two conditions. Children participated in four familiarization trials followed by four disambiguation trials.

Table 2
Pairs of Familiar–Novel Objects Used in Study 4, and the New Functions Requested in the Disambiguation Phase

Familiar object	Novel object	New function
Food plate	Round wooden block	Make noise
Toy hat	Oddly shaped sponge	Wipe off sand
Plastic bottle	Ceramic cone	Smoosh dough
Color marker	Dumpling maker	Make holes

Procedure

The procedure in Study 4 was nearly identical to the procedure in Study 1. The only difference was that in Study 4 children were familiarized with a user who was either knowledgeable or ignorant of objects' functions, and who later asked children for the object needed to perform a new function.

Familiarization phase. The familiarization phase was the only phase that differed between the two conditions. In this phase, the experimenter invited the child and the puppet to play, and presented familiar objects one at a time. The puppet looked at the object and stated what the object was used for three times while asserting his knowledge of what the object is (i.e., saying, "I know what this is!"). The main manipulation had to do with whether or not the functions mentioned by the puppet were the conventional functions associated with the objects. In the knowledgeable condition, the puppet mentioned the conventional functions of the objects (e.g., upon seeing a spoon, he said repeatedly, "That's for eating!"). In the ignorant condition; the puppet mentioned incorrect functions for the objects (e.g., about the spoon, "That's for cutting paper!"). After a few seconds in which the child and the puppet were allowed to play with an object, the experimenter returned the object to her bag and brought out another familiar object. Again the puppet reacted according to the child's condition. This procedure was repeated for four trials in order to allow the child to build an assessment of the puppet's knowledge—or ignorance—of familiar objects' functions. The conventional functions used for the items, chair, spoon, hammer, and comb, were "for sitting," "for eating," "for pounding," and "for combing hair," respectively. The incorrect functions used were "for sweeping floors," "for cutting paper," "for singing," and "for picking flowers," respectively. Order of presentation of the objects was randomized across children, as in Study 1.

Disambiguation phase. This phase was identical in both conditions. The experimenter laid down in front of the child the first pair of familiar–novel objects. The puppet then looked at the child, and asked for one of the objects by using a new function that could be performed with either object (e.g., "Can you give me the one *for making noise*?"). The functions associated with each object pair were the same for all children. This was done because the functions were created so as to be affordable by both objects in each pair. The new functions applied to each pair are listed in Table 2. Both the puppet and the experimenter refrained from looking at either one of the objects and instead maintained their gaze on the child. Once the child responded by pointing to, picking up, or handing to the puppet one of the objects, the experimenter removed the two objects and replaced them with the next pair of familiar–novel objects. The experimenter did not provide any feedback to children about their performance and instead made comments about the game and the toys. The experimenter wrote down children's choices and any other verbal comments children made throughout the entire procedure. Upon completion of the four disambiguation trials, children were thanked, given a sticker in appreciation for their participation, and returned to their class.

The order of presentation of the pairs was counterbalanced across children, as in Study 1. The left–right placement of each object within a pair was counterbalanced within children, so that on half of the pairs, the familiar object was placed on the right side of the child, and on the other half of the pairs, the novel object was placed on the right side of the child.

Results

Our main prediction was that children would be more likely to choose the novel object in response to a request made by a knowledgeable user than one made by an ignorant user. The dependent measure used to address this prediction was the number of times (0–4) children selected the novel object as the referent of the new function requested by the puppet in the disambiguation phase.

An ANOVA with condition (knowledgeable vs. ignorant) and gender as between-subjects variables, and participants' age in months as a covariate, on the above dependent measure revealed a main effect of condition, $F(1, 31) = 4.53$, $p < .05$, $\eta_p^2 = .13$; no effect of gender, $p > .6$; no effect of age, $p > .7$; and no interaction between gender and condition, $p > .9$. As illustrated in Figure 1, children in the knowledgeable condition ($M = 3.56$, $SD = 0.70$) were more likely than children in the ignorant condition ($M = 2.94$, $SD = 0.94$) to pick the novel object in response to the puppet's request for a new function. As in Study 1, analyses against chance (chance = 2) revealed that children in both conditions were more likely to pick the novel object than would be expected by chance: in the knowledgeable condition, $t(17) = 9.36$, $p < .001$, in the ignorant condition, $t(17) = 4.27$, $p < .005$. Finally, the interaction between age as a dichotomous variable based on a median split of the sample, and condition, was not significant, $p > .8$.

In order to evaluate children's individual pattern of responses, we conducted the same nonparametric test run in Study 1, evaluating the distribution of children in terms of response consistency. As shown in Table 1, while the distribution pattern in Study 4 was similar to that found in Study 1, the difference between conditions in Study 4 did not reach significance, $\chi^2(1, N = 36) = 2.57$, $p = .11$.

Taken together, the results of Study 4 were consistent with our hypothesis. Children were more likely to select a novel object as the referent of a new function when the user asking for the function had previously manifested knowledge of the conventional functions of objects than when the user manifested ignorance of conventional functions.

Comparisons Between Studies 1 and 4

Given that Studies 1 and 4 had the same design and procedure, a comparison between them allowed us to directly assess potential differences between children's responses to labels (Study 1) and functions (Study 4). We conducted an ANOVA using Study (1 vs. 4) and condition (knowledgeable vs. ignorant) as between-subjects variables on the mean number of times children selected the novel object in response to the puppet's request (for either a novel label, in Study 1, or a new function, in Study 4). The ANOVA revealed a significant effect of condition, $F(1, 65) = 10.71$, $p < .005$, $\eta_p^2 = .14$; no effect of study, $p > .5$; and no interaction, $p > .7$. These findings were confirmed when looking at the distribution of children in terms of disambiguators or overlap accepters across these two studies. In the knowledgeable conditions across the two studies, 32 of the 34 children were disambiguators, whereas in the ignorant conditions, only 22 of the 35 were so, $\chi^2(1, N = 69) = 9.91$, $p < .005$. In terms of studies, 26 of the 33 children in Study 1 were disambiguators, compared to 28 of the 36 in Study 4, evidently a nonsignificant difference, $p > .9$.

Altogether, the knowledge state of the requester affected children's responses in the same way, whether he was asking for the referents of labels or for the functions of objects. Moreover, children did not show a stronger tendency to avoid overlap between two labels than between two functions.

Discussion

The goal of Study 4 was to assess whether children's inferences about the referents of functions would be similar to their inferences about the referents of object names. The findings indicate that they were. Children in Study 4 were more likely to infer a contrastive intent when a user seemed knowledgeable about the conventional functions of objects than when a user seemed ignorant of such conventions. In fact, the direct comparison of children's responses to object functions and their responses to object labels revealed no significant differences. Given the similarity in the methodologies used to assess children's responses in these two domains, the above findings intimate that the same inferential process underlies children's responses in both domains.

General Discussion

A fundamental assumption for successful communication to take place is that partners in a conversation share knowledge about the expected ways to refer to familiar concepts. This assumption of conventionality has been documented in children, primarily via its contrastive implication, namely, that if a speaker uses a form that is different from the one expected in a given context, then that speaker likely has a different concept in mind (Clark, 1988). A number of studies have shown that children indeed draw this kind of inference when a conventional *form* is expected but is not used (Diesendruck, 2005; Diesendruck & Markson, 2001; Henderson & Graham, 2005). One of the central goals of the present studies was to address whether the extent to which a *source* has proven to be knowledgeable of conventional forms affects children's tendency to infer contrastive intents.

The results of the present studies reveal that it does. In Study 1, when a speaker proved knowledgeable of familiar names, then his subsequent request for the referent of a novel label was taken by children as indicative of an intent to refer to a novel—rather than a familiar—object. Children drew this inference to a significantly lesser extent when the speaker had manifested ignorance of the conventional names of familiar objects. Study 2 confirmed these findings and provided evidence that children's inferences, particularly those about an ignorant speaker's requests, were not driven by considerations of playfulness or silliness but rather considerations of knowledge states (see also Koenig & Harris, 2005, on this point).

A social-pragmatic account of these findings is that children expected a knowledgeable speaker to know the names of familiar objects and expected such a speaker to use these conventional names if intending to refer to the familiar objects. Given that the speaker used novel names, children inferred that the speaker had a contrastive intent in mind. In turn, given that the ignorant speaker proved not to know the conventional names of familiar objects, children relaxed their expectation of conventionality toward him and thus did not as decisively draw a contrastive inference in response to his use of a novel label. The present findings cannot determine whether evidence for knowledge reinforced an assumption of conventionality, evidence for ignorance weakened it, or both. What is crucial for the present purposes, however, is that children modulated their contrastive inferences based on their assessments of speakers' knowledge states.

A previous study (Diesendruck, 2005) had shown that children withhold an assumption of conventionality from a speaker who does not speak the language in which children were taught a novel name. The present findings add to that by revealing that children's modulation of an assumption of conventionality operates even within a single language and may lead children to doubt that a speaker knows even the names of highly familiar objects. On a separate issue, these findings add to a growing number of studies indicating that children's avoidance of lexical overlap is not driven by lexical constraints (Golinkoff et al., 1994; Markman & Wachtel, 1988; Merriman & Bowman, 1989). Rather, children are willing to accept two names for an object, especially in cases where an assumption of conventionality regarding the speakers cannot be upheld (Diesendruck, 2005; Sabbagh, Wdowiak, & Ottaway, 2003).

The primary goal of Study 3 was to assess the leniency of children's assumption of conventionality. As Clark (1997) has advocated, an assumption of conventionality has to be flexible enough to allow speakers to deviate from rigid expectations of what should be said without treating them as outcasts in one's linguistic community whose knowledge of the language cannot be trusted. The results of Study 3 confirmed that children respond to a speaker who refers to familiar objects with peculiar yet linguistically correct labels in the same way in which they respond to a speaker who uses conventional familiar labels to name those objects. In other words, the peculiarity in the labels used by Study 3's speaker to refer to familiar objects did not throw children off and did not lead them to withhold an assumption of conventionality towards that speaker.

A comparison of these results to those in Studies 1 and 2 ignorant conditions, and those reported by Diesendruck (2005), intimates that children are quite resistant to abandoning an assumption of conventionality. In particular, children seem to uphold an assumption of conventionality toward an otherwise mature speaker of a language who momentarily mislabeled familiar objects (ignorant condition of Study 1) or who provided peculiar but correct labels to familiar objects (unusual condition of Study 3). In turn, children withhold an assumption of conventionality toward an immature speaker of a language who momentarily mislabeled familiar objects (ignorant condition of Study 2) or a speaker who does not know the language at all (Diesendruck, 2005). In sum, children might operate with the default assumption that mature speakers of the child's native language are knowledgeable of the conventional forms prevalent in the language.

It is important to point out that the above conclusion may be true only of the particular aspect of conventionality assessed in the present studies. That is, children may have a lenient assumption that speakers ignorant of certain familiar object labels nonetheless know the names of other *familiar* objects in their language. It is possible, nonetheless, that children would treat such speakers as untrustworthy teachers of *novel* labels and thus would not treat novel labels used by them as conventional. A few studies are consistent with this possibility (Birch et al., 2008; Buresh & Woodward, 2007; Graham et al., 2006), but the issue of leniency in light of atypical labeling has not been directly addressed in this regard.

Study 4 evaluated whether children respond in a similar way to users who violate the conventional functions of objects as they do to speakers who violate the conventional names of object. The findings from that study confirmed that children indeed treat these two cases equivalently. Children in Study 4 were just as sensitive to the knowledge state of object *users* when inferring which object a user had in mind for performing a *function*, as they were to the knowledge state of object *labelers* when inferring which object a speaker had in mind as the referent of a *label*. Namely, they drew contrastive inferences when a user was knowledgeable more often than when the user was ignorant. It is important to note that while in the present study knowledge of function was instantiated by the puppet *naming* the function—as opposed to actually *performing* it—studies by Casler and Kelemen (2005, 2007) found that children draw contrastive inferences also in cases where the functions are performed. In other words, we believe that the patterns revealed in the present study are indeed representative of how children reason about the mapping of functions to objects.

An additional finding from Study 4 was that, just as in the case of label, children manifested some resistance to abandoning an assumption of conventionality toward object functions altogether. Specifically, even children in the ignorant condition of Study 4 selected the novel object as appropriate for performing a new function more than would be expected by chance. Again, it is as if children reasoned that while the user might not know what the specific familiar objects in the familiarization phase are for (e.g., saying that a spoon is for cutting), he must know, *in general*, what familiar objects are for (e.g., that a plate is not for making noise). Thus, momentary evidence for ignorance of conventions merely weakens children's default assumption in this case as well.

The idea that a reasoning process that invokes an assumption of conventionality and inferences about users' intentions underlies children's assignment of functions to objects is consistent with a number of recent findings regarding artifact categories in children. First, this conclusion is consistent with the notion that children's concepts of artifacts are informed by beliefs regarding the artifacts' intended function (Bloom, 1996; Diesendruck, Markson, & Bloom, 2003). Children do not read off functions from objects directly but rather via guessing what somebody meant the object to be. Second and more central to the present proposal, the conclusion is consistent with findings that by 2½ years of age, children believe that objects have unique purposes, which are known by all members of a culture (Casler & Kelemen, 2005). Children not only expect others to know what objects are for, but, as the present findings indicate, they draw contrastive inferences about users' intended function based on their assessment of users' knowledge of conventional functions.

A crucial question this conclusion invites is what is the scope of an assumption of conventionality? One way to address this question is by evaluating the processes by which children come to develop an assumption of conventionality. According to one proposal, conventionality results from children's limited capacity to entertain disparities in the knowledge states of different individuals (Sabbagh & Henderson, 2007). In other words, conventionality is a manifestation of children's overgeneralization of their knowledge states onto others—that is, if they know something, then others probably know it as well (see also Birch & Bloom, 2003, on this so-called curse of knowledge). This proposal seems to imply that children's assumption of conventionality is fairly general in application. Challenging this proposal are findings that children *do* modulate their assumption of conventionality depending on others' knowledge state (see present studies; Diesendruck, 2005; Sabbagh et al., 2003) and on the specific forms used (e.g., the cases of proper names, Diesendruck, 2005; and facts, Diesendruck & Markson, 2001). In fact, recent studies reveal that even prior to their second birthday, infants expect object names to be shared by distinct individuals but do not hold such expectation regarding preferences (Graham et al., 2006) or goals (Buresh & Woodward, 2007). In other words, there are limits to the scope of children's assumption of conventionality.

A second proposal is that an assumption of conventionality derives from parental input. In particular, Callanan and colleagues have suggested that the way in which parents label objects when speaking to children (Callanan & Sabbagh, 2004) and the way in which parents demonstrate the uses of objects to children (Callanan, Siegel, & Luce, 2007) may endorse the notion that there is *a best way* to label and use objects. According to this proposal, conventionality would be limited to knowledge forms about which children receive this type of rich parental input.

A third proposal for the origin of conventionality with implications regarding its scope, focuses on cognitive capacities available to children from early in development. Two capacities, in particular, seem relevant to the development of an assumption of conventionality: (a) children's understanding of intentions, as already hinted above, and (b) children's capacity to detect statistical regularities in the input. Notice that while compatible with an input account, this proposal emphasizes what the child brings to the task of inferring conventionality, rather than how the input may be so structured as to facilitate this inference.

There is a vast literature indicating that already prior to their second birthday, children interpret people's actions as intentional or goal directed (Gergely, Nadasdy, Csibra, & Biro, 1995; Meltzoff, 1995; Woodward, 1998). It has further been found that an understanding of intentions is involved in young children's inferences about word meanings (Akhtar & Tomasello, 2000; Baldwin, 1991; Bloom, 2000) and, as noted above, artifact concepts (Bloom, 1996). The proposal is that children's monitoring of adults' actions in terms of intentionality might help them determine which "human products" are conventional and which are not. Evidently, this would not be sufficient for such a determination, but it might nonetheless be necessary.

In terms of children's statistical learning capacity, it has been shown that even prior to their first birthday infants are capable of picking out patterns in a stream of stimuli (Saffran, 2003). The idea being put forth here is that detection of stability in the way people refer to or use objects may contribute to children's assumption that an object label or function is conventional. In fact, Siegel and Callanan (2007) found that children's determination of what an object was for was influenced by the number of people said to use the object in a specific way.

One of the implications of grounding an assumption of conventionality in terms of broader cognitive capacities is that we can then draw on the accumulating knowledge about the development of these capacities to explain, and generate hypotheses about, patterns regarding the development of conventionality. For instance, in none of the studies reported here was there an effect of age on children's performance, nor an interaction between age and experimental condition. That is, children within a broad range of ages between 3 and 4 years of age responded similarly in our tasks. In contrast to this null result, there are inconsistencies with regard to the effect of age among studies assessing children's willingness to learn new information based on speakers' reliability. A possible reconciliation of these inconsistencies has to do with the nature of the tasks. In particular, there do not appear to be age differences in tasks that require children to rely primarily on their assessments of others' knowledge in order to infer others' intended behavior (e.g., the present studies; Birch et al., 2008; Jaswal & Neely, 2006; Koenig et al., 2004; explicit judgment task). In turn, age differences were found in tasks that required children to predict a speaker's future behavior (e.g., Koenig & Harris, 2005) or in which children had to make subtle calculations about a speaker's state of knowledge (Pasquini, Corriveau, Koenig, & Harris, 2007). In a sense, this distinction seems analogous to ones made in the theory-of-mind literature between explaining behavior in terms of mental states versus predicting behavior in terms of mental states (e.g., Bartsch & Wellman, 1989).

The conclusion regarding conventionality is that one of the crucial capacities that undergo development in the preschool years is not so much the capacity to *assess* others' state of knowledge but rather the capacity to *apply* productively and originally that assessment in pertinent real world situations. Consistent with this conclusion, recent studies show that even prior to their second birthday, infants may be capable of assessing others' knowledge states (Onishi & Baillargeon, 2005), interpret people's behavior in terms of knowledge states (Moll, Koring, Carpenter, & Tomasello, 2006), and expect others to share certain knowledge states (Buresh & Woodward, 2007). It would be valuable for future studies to develop paradigms in order to assess children's capacity to apply conventionality flexibly.

A further implication of this latter proposal is that the knowledge forms that are treated by children as conventional are those that, while arbitrary and thus without a transparent meaning, nonetheless tend to be intentionally produced and stably represented. By many accounts, these characteristics are precisely what define culture (e.g., Csibra & Gergely, 2006; Tomasello, Carpenter, Call, Behne, & Moll, 2005). Thus, one can speculate that other cultural forms, such as symbolic representations, norms, or communicative signs, will also be treated as conventional by children. In this sense, the research program that investigates the development of children's assumption of conventionality meshes with the program investigating children's acquisition of culture.

References

Akhtar, N., Carpenter, M., & Tomasello, M. (1996). The role of discourse novelty in early word learning. *Child Development, 67*, 635–645.

Akhtar, N., & Tomasello, M. (2000). The social nature of words. In R. M. Golinkoff & K. Hirsh-Pasek (Eds.), *Becoming a word learner: A debate on lexical acquisition* (pp. 115–135). New York: Oxford University Press.

Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development, 62*, 875–890.

Bartsch, K., & Wellman, H. (1989). Young children's attribution of action to beliefs and desires. *Child Development, 60*, 946–964.

Birch, S. A. J., & Bloom, P. (2002). Preschoolers are sensitive to the speaker's knowledge when learning proper names. *Child Development, 73*, 434–444.

Birch, S. A. J., & Bloom, P. (2003). Children are cursed: An asymmetric bias in mental-state attribution. *Psychological Science, 14*, 283–286.

Birch, S. A. J., Vauthier, S. A., & Bloom, P. (2008). Three- and four-year-olds spontaneously use others' past performance to guide their learning. *Cognition, 107*, 1018–1034.

Bloom, P. (1996). Intention, history, and artifact concepts. *Cognition, 60*, 1–29.

Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.

Buresh, J. S., & Woodward, A. L. (2007). Infants track action goals within and across agents. *Cognition, 104*, 287–314.

Callanan, M. A. (1985). How parents name objects for young children: The role of input in the acquisition of category hierarchies. *Child Development, 56*, 508–523.

Callanan, M. A., & Sabbagh, M. A. (2004). Multiple labels for objects in conversations with young children: Parents' language and children's developing expectations about word meanings. *Developmental Psychology, 40*, 746–763.

Callanan, M. A., Siegel, D. R., & Luce, M. R. (2007). Conventionality in family conversations about everyday objects. *New Directions for Child and Adolescent Development, 115*, 83–98.

Casler, K., & Kelemen, D. (2005). Young children's rapid learning about artifacts. *Developmental Science, 8*, 472–480.

Casler, K., & Kelemen, D. (2007). Reasoning about artifacts at 24 months: The developing teleo-functional stance. *Cognition, 103*, 120–130.

Clark, E. V. (1988). On the logic of contrast. *Journal of Child Language, 15*, 317–335.

Clark, E. V. (1990). On the pragmatics of contrast. *Journal of Child Language, 17*, 417–431.

Clark, E. V. (1997). Conceptual perspectives and lexical choice in acquisition. *Cognition, 64*, 1–37.

Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In M. H. Johnson & Y. Munakata (Eds.), *Processes of change in brain and cognitive development, attention, and performance* (pp. 249–274). Oxford, UK: Oxford University Press.

Diesendruck, G. (2005). The principles of conventionality and contrast in word learning: An empirical examination. *Developmental Psychology, 41*, 451–463.

Diesendruck, G., & Markson, L. (2001). Children's avoidance of lexical overlap: A pragmatic account. *Developmental Psychology, 37*, 630–641.

Diesendruck, G., Markson, L., Akhtar, N., & Reudor, A. (2004). Two-year-olds' sensitivity to speakers' intent: An alternative account of Samuelson and Smith. *Developmental Science, 7*, 33–41.

Diesendruck, G., Markson, L., & Bloom, P. (2003). Children's reliance on creator's intent in extending names for artifacts. *Psychological Science, 14*, 164–168.

Gergely, G., Nadasdy, Z., Csibra, G., & Biro, S. (1995). Taking the intentional stance at 12 months of age. *Cognition, 56*, 165–193.

German, T. P., Truxaw, D., & Defeyter, M. A. (2007). The role of information about "convention," "design," and "goal" in representing artificial kinds. *New Directions for Child and Adolescent Development, 115*, 69–82.

Golinkoff, R. M., Mervis, C. B., & Hirsh-Pasek, K. (1994). Early object labels: The case for a developmental lexical principles framework. *Journal of Child Language, 21*, 125–156.

Graham, S. A., Stock, H., & Henderson, A. M. E. (2006). Nineteen-month-olds' understanding of the conventionality of object labels versus desires. *Infancy, 9*, 341–350.

Hall, D. G. (1994). Semantic constraints on word learning: Proper names and adjectives. *Child Development, 65*, 1299–1317.

Henderson, A. M. E., & Graham, S. A. (2005). Two-year-olds' appreciation of the shared nature of novel objects labels. *Journal of Cognition and Development, 6*, 381–402.

Jaswal, V. K., & Neely, L. A. (2006). Adults don't always know best: Preschoolers use past reliability over age when learning new words. *Psychological Science, 9*, 757–758.

Kalish, C. W., & Sabbagh, M. A. (2007). Conventionality and cognitive development: Learning to think the right way. *New Directions for Child and Adolescent Development, 115*, 1–10.

Koenig, M. A., Clement, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science, 15*, 694–698.

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development, 76*, 1261–1277.

Lampinen, J. M., & Smith, V. L. (1995). The incredible (and sometimes incredulous) child witness: Child eyewitnesses' sensitivity to source credibility cues. *Journal of Applied Psychology, 80*, 621–627.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meaning of words. *Cognitive Psychology, 20*, 121–157.

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology, 31*, 838–850.

Merriman, W. E., & Bowman, L. L. (1989). The mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child Development, 54*(3–4, Serial No. 220).

Moll, H., Koring, C., Carpenter, M., & Tomasello, M. (2006). Infants determine others' focus of attention by pragmatics and exclusion. *Journal of Cognition and Development, 7*, 411–430.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*, 255–258.

Pasquini, E. S., Corriveau, K. H., Koenig, M., & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. *Developmental Psychology, 43*, 1216–1226.

Rakoczy, H. (2007). Play, games, and the development of collective intentionality. *New Directions for Child and Adolescent Development, 115*, 53–68.

Sabbagh, M. A., & Baldwin, D. A. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschoolers' theory of mind and semantic development. *Child Development, 72*, 1054–1070.

Sabbagh, M. A., & Henderson, A. M. E. (2007). How an appreciation of conventionality shapes early word learning. *New Directions for Child and Adolescent Development, 115*, 25–38.

Sabbagh, M. A., Wdowiak, S. D., & Ottaway, J. M. (2003). Do word learners ignore ignorant speakers? *Journal of Child Language, 30*, 905–924.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science, 12*, 110–114.

Siegel, D. R., & Callanan, M. (2007). Artifacts as conventional objects. *Journal of Cognition and Development, 8*, 182–203.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences, 28*, 675–691.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69*, 1–34.

Child Development, March/April 2010, Volume 81, Number 2, Pages 669–679

Children Monitor Individuals' Expertise for Word Learning

David M. Sobel
Brown University

Kathleen H. Corriveau
Harvard University

Two experiments examined preschoolers' ability to learn novel words using others' expertise about objects' nonobvious properties. In Experiment 1, 4-year-olds (n = 24) endorsed individuals' labels for objects based on their differing causal knowledge about those objects. Experiment 2 examined the robustness of this inference and its development. Four-year-olds (n = 40) endorsed labels from confederates who accurately predicted objects' nonobvious internal properties but not nonobvious external properties. Three-year-olds (n = 40) performed at chance levels in both cases and were less likely to recognize the informants' expertise, suggesting that they might be unable to monitor individuals' expertise. These data suggest that children's ability to learn from testimony is necessary for their understanding of the relevance of an individual's expertise.

As adults, we rely on others' knowledge to learn and to make inferences about the world. We count on mechanics to diagnose and repair our cars because they know more about the functions of fuel-injection systems and carburetors. We allow doctors to advise us on medical decisions because they know more about how diseases cause symptoms and are affected by treatments. We hope our political leaders can effect change in social issues, the economy, or foreign affairs because of their greater experience or deeper knowledge of the issues involved. Whether trust is warranted in a mechanic, doctor, or politician depends on our belief that they understand the causal structures critical to a specific knowledge base. This article concerns whether and at what age children can use an individual's causal expertise to make novel inferences. Specifically, we focus on the effect observing another's causal knowledge has on children's subsequent willingness to rely on that person when learning new words.

The importance of "trust" in word learning has been studied extensively because there is a conventional, but arbitrary, assignment of lexical items to concepts and objects (e.g., Bloom, 2000; de Saussure, 1966; see Koenig & Harris, 2008, for a review). As such, children must rely on data they observe to learn the meaning of words. This involves recognizing that others are reliable sources of information. Thus, children should use data gained from a reliable source while discounting information that is unreliable. A growing body of research indicates that by the age of 4 (and possibly younger), preschoolers are able to use a speaker's past history of reliability to make inferences about the extension of novel labels to novel objects (e.g., Clément, Koenig, & Harris, 2004; Corriveau & Harris, 2009; Jaswal & Neely, 2006; Koenig, Clément, & Harris, 2004; Koenig & Harris, 2005; Pasquini, Corriveau, Koenig, & Harris, 2007). Four-year-olds also assume that adults tend to be reliable by default (e.g., Corriveau, Meints, & Harris, 2009) as well as register the role of uncertainty in word learning (e.g., Sabbagh & Baldwin, 2001). These data all suggest that children trust adults to provide them with accurate lexical information and discount that information when the adult is a less reliable source (but see Nurmsoo & Robinson, 2009, for a different opinion).

Such inferential ability potentially allows children to develop concepts of expertise. Children develop ontological commitments about other kinds of events they cannot directly observe based on the "patterns of testimony" that they receive (Harris & Koenig, 2006; Harris, Pasquini, Duke, Asscher, & Pons, 2006). More generally, Keil and colleagues have suggested that young children learn that individuals' knowledge forms a "division

---

This research was supported by NSF (DLS-0518161 to DMS). We would like to thank Sheridan Brett, David Buchanan, Silvia Chen, Jon Cohen, Angie Kim, Karina Ikesoe, Emily Myers, Rachel Shelley-Abrahamson, Cesalie Stepney, Elena Tenenbaum, and Lea Ventura for help with data collection and analysis.

Correspondence concerning this article should be addressed to David M. Sobel, Department of Cognitive and Linguistic Sciences, Box 1978, Brown University, Providence, RI 02912. Electronic mail may be sent to dave_sobel@brown.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8102-0019

of cognitive labor'' (Danovitch & Keil, 2004; Keil, Stein, Webb, Billings, & Rozenblit, 2008; Lutz & Keil, 2002). At relatively young ages, children understand that individuals possess different pieces of knowledge, which potentially cohere around causal principles. Keil et al. (2008), in particular, demonstrated that 5-year-olds recognize that others' knowledge is organized around disciplines that reflect the structure of modern universities. Children understood that someone who knows about a physical property is more likely to know about a chemical one (grouping two natural sciences) than a principle of cognitive psychology (grouping a natural science with a social science).

Understanding this ''division of cognitive labor'' requires that children have access to and can make inferences about their own knowledge. Our goal here is not to investigate the emergence of this metacognitive understanding but rather to consider whether children possess the cognitive ability to monitor individuals as reliable sources of knowledge within a domain. This might be the basis on which the metacognitive division of cognitive labor is built. In particular, we are interested in the consistency between individuals' expertise and the inference children are asked to make based on that expertise. For example, several researchers have found that preschoolers rely on speakers' history of accurately labeling familiar objects when those speakers subsequently provided conflicting information about a novel object's function (Birch, Vauthier, & Bloom, 2008; Koenig & Harris, 2005). Using the more reliable speaker in this circumstance seems logical given the coherence between artifacts' category membership and their functions. Preschoolers often relate these two pieces of information, making inferences about one based on the other (e.g., Bloom, 1996; Gelman & Bloom, 2000; Kemler Nelson, Frankenfield, Morris, & Blair, 2000; Kemler Nelson et al., 1995). The reliability with which an individual categorizes objects should inform children that the individual possesses categorical knowledge about those objects, such as their function. But an individual's history of reliabilty should only be applicable if that reliability is relevant to the inference at hand—as reliable as a plumber might be at fixing a toilet, we might choose to discount his political savvy.

These insights motivated this study. In Experiment 1, we examined whether 4-year-olds could recognize that two individuals were experts in different pieces of causal knowledge. We asked whether 4-year-olds could subsequently use the individuals' expertise when learning the label of a novel object that exhibited a causal property related to only one of the individual's expertise. Several investigations have shown that young children relate objects' causal properties with their category membership, particularly when assigning labels to those objects (Booth & Waxman, 2002; Gopnik & Sobel, 2000; Graham, Kilbreath, & Welder, 2004; Nazzi & Gopnik, 2000). Although some have argued that object labels do not necessarily relate to conceptually deep properties of those objects, like their kind or function (e.g., Smith, Jones, & Landau, 1996), we suggest that children recognize that object labels relate to the conceptual structure indicated by those labels. Thus, we hypothesized that young children would rely on individuals' different causal knowledge judiciously to make inferences about novel objects' labels.

Experiment 2 had three related goals. The first was to examine whether preschoolers used informants' reliability about a different property of objects—its insides or material kind—to make inferences about novel object labels. Several investigations have demonstrated that 4-year-olds, but not 3-year-olds, understand that objects' causal properties or category membership are related to those objects' insides (Gelman & Gottfried, 1996; Gelman & Wellman, 1991; Gottfried & Gelman, 2005; Sobel, Yoachim, Gopnik, Meltzoff, & Blumenthal, 2007). Given this developmental difference, we hypothesized that 4-year-olds, but not 3-year-olds, would relate individuals' accurate histories of describing nonobvious internal properties of objects with their category membership.

But, an individual's reliabilty should only be applicable if that reliability is relevant to the inference at hand. The second goal of Experiment 2 was to consider whether children treated any set of reliable actions as relevant to word learning or whether children distinguished between relevant and irrelevant training. To investigate this possibility, we presented another group of 3- and 4-year-olds with informants who possessed different knowledge about an external property of a set of objects. Half of the objects had green stickers on them and half had red stickers. One informant was accurate in identifying which objects had red stickers on them but was ignorant about the other objects; the other informant was accurate in identifying the objects with green stickers on them but was ignorant about the objects with red stickers. Our hypothesis was that this expertise would be irrelevant to the objects' category membership: Neither age group should use the speakers' reliability about these external parts to make inferences about category membership. Furthermore, the 4-year-olds in particular should be more likely to use the speakers' reliability in the insides condition than in the outsides condition.

Experiment 1

Four-year-olds were introduced to two individuals who demonstrated different pieces of causal knowledge about the way in which objects activated a machine. We then showed children a new object whose efficacy on the machine was consistent with the predictions of one of the two individuals. We examined whether children would use the individuals' different knowledge to decide which informant to turn to when making an inference about the label of the new object.

Method

Participants

Twenty-four 4-year-olds (16 boys and 8 girls, $M = 53.71$ months, $SD = 2.90$) were recruited from a preschool classroom, fliers posted at local preschools, and a list of hospital births. Four additional children were tested but not included in the final analysis: 2 because of experimental error and 2 failed to interact with the experimenter. Most children were Caucasian and from middle- to upper-middle-class families; however, no formal measure of socioeconomic status (SES) was taken. The participants did represent the diversity of the population from which they were sampled.

Materials

The machine was an 8 × 6 × 3 in. box made of black plastic with a white top (see Figure 1). If an object was placed on it, the machine would activate, both by illuminating either red or green light-emitting diodes (LEDs), which were wired into the top of the device, and by playing one of two different musical arrangements. The machine worked via a remote control (hidden from the child), which the experimenter used to control whether the machine activated red or green. All of the objects used activated the machine, but the remote could also "enable" or "disable" the machine, such that other objects that a child might inadvertently put on the machine (e.g., fingers) would not activate it. Eight wooden blocks, all of different sizes, colors, and shapes were used. No block was red or green.

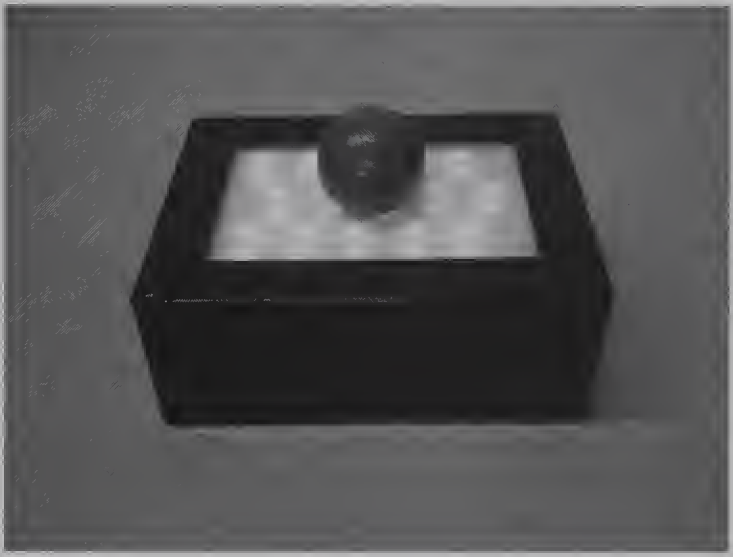


Figure 1. The machine used in the Experiment 1.
Note. The machine would activate by illuminating red or green light-emitting diodes (LEDs) wired into the white top. When the machine activated, it also played one of two songs, dependent on whether it was set to light up red or green. The objects used in Experiment 1 were all wooden blocks without any distinguishable features. The object shown in Figure 1 is an object used in Experiment 2, with a dowel over a cavity, revealing its "inside" (in the insides condition) and with a sticker on the back (not shown), both of which were either red or green.

Procedure

Children were tested in a quiet room at their school or in the lab. The child and experimenter were seated across from each other at a table. The experimenter told the child that they would play a game with two of the experimenter's friends. The two confederates then entered the room and sat on either side of the experimenter. The confederates were always the same gender and never wore red or green clothing. The child was then told that they would all play a game together with the experimenter's machine, which was brought out onto the table. The experimenter said, "This is a special machine. I have some toys. Some toys make the machine go red and some toys make the machine go green, and we're going to try and figure it out. Let's take a look."

Training trials. Children observed six training trials. In each, the experimenter brought out a block (randomly chosen from the set of eight) from under the table. On the first training trial, the experimenter told the child, "Take a look at this toy. I don't know what's going to happen when we put this on the machine, but my friends are pretty smart, so let's ask them." (On subsequent trials, the experimenter just showed the child the new object and said, "Let's ask my friends what this one does.")

The experimenter turned to one of the confederates (counterbalanced across children) and asked "What do you think will happen when we put this on the machine?" Each confederate was given a chance to respond. Prior to the start of the procedure (and unbeknownst to the child), one confederate was designated the *red expert* whereas the other confederate was the *green expert*; we randomly determined who the red expert was for each participant. The red expert always predicted when an object would activate the machine red but was ignorant of objects that made the machine go green (and responded "I don't know" for those objects). Similarly, the green expert always responded correctly for objects that made the machine go green and was ignorant for objects that made the machine go red.

After each confederate stated her belief about what the object would do, the experimenter repeated their claims and then placed the object on the machine, which activated consistent with the expert's claim. Children were then asked to point to the person who was right (i.e., who correctly predicted the event). Corrective feedback was given if necessary. These questions were included as a catch trial—our intention was to replace the data from any child who answered more than one of these questions incorrectly from the final analysis. However, all children answered this question correctly on at least five of the six trials. On three of the six training trials, the object activated the machine red and, on the other three, the object activated the machine green (order was randomly determined). Thus, children saw three trials in which the red expert was accurate and the green expert was ignorant (all about red activation events), and three trials in which the green expert was accurate and the red expert was ignorant (all about green activation events). Children never observed either confederate respond incorrectly.

Test trials. Immediately after the training trials, children then observed two test trials. In each, the child was shown a new object and was told "Let's see what this one does on the machine." The object was placed on the machine without asking either confederate her belief about its efficacy. On one trial, the object made the machine go red; on the other, it made the machine go green. The experimenter narrated the machine's activation for the child (e.g., "Look, this one made the machine go green"). The experimenter then said to the child, "Do you see this toy. I'll tell you a secret. This toy has a special name. I don't know the special name of this toy, but we can ask my friends." Children were then asked the *ask* question, "Who do you want to ask for the special name of the toy, [Confederate 1] or [Confederate 2]?"

After the child responded, the experimenter asked each confederate for the special name of the toy. Each confederate generated a novel label (*dax* vs. *wug* on one trial, *modi* vs. *toma* on the other). Each label was repeated for the child, and then the child was asked the *endorse* question—for example, "What do you think the special name of this toy is called, a *modi* or a *toma*?"

After these two trials, children were asked two *knowledge* questions, "Who knows about things that make the machine go red?" and "Who knows about things that make the machine go green?" These questions were asked to ensure that children remembered the expertise of the two confederates. The location of the red and green experts (with respect to the experimenter) and the particular word labels the experts generated were both counterbalanced.

Results

Overall, children required feedback on ~4% of the training trials, and 18 of the 24 children required no feedback at all. This suggested that children understood the training. Children received a score of 1 on each question if they responded consistently with the confederate with the relevant causal expertise (e.g., if the test object activated the machine green, children were given a score of 1 if they asked the green expert for the special name of the toy on the ask question and endorsed that confederate's label on the endorse question). Preliminary analyses suggest that these scores did not differ between the two ask questions, the two endorse questions, or the two knowledge questions, all McNemar $\chi^2(1, n = 24)$ tests = *ns*, so we combined them. These data are shown in Table 1.

We first analyzed responses on the ask and endorse questions. Only responses on the endorse questions showed a pattern significantly different from chance levels, $\chi^2(2, n = 24) = 6.92$, $p = .031$. The distribution of responses to the ask question was not different from chance, $\chi^2(2, n = 24) = 0.92$, *ns*. Responses to neither of these questions significantly correlated with age.

We next examined responses to the knowledge question. The distribution of responses on this question was significantly different from chance,

Table 1
Distribution of Responses to the Ask, Endorse, and Knowledge Questions in Experiment 1

	Number of correct responses				
	0	1	2	*M*	*SD*
Question type					
Ask	5	11	8	1.12	0.74
Endorse	2	11	11	1.38	0.65
Knowledge	1	2	21	1.83	0.48

with 21 out of the 24 children responding correctly on both trials, $\chi^2(2, n = 24) = 50.00$, $p < .001$. As with the ask and endorse questions, responses did not correlate with the child's age.

Finally, we wanted to be certain that responses to the ask and endorse questions were representative of children who understood the expertise of the informants. We reanalyzed these responses looking only at the children who answered both knowledge questions correctly. These analyses paralleled what was presented earlier: On the endorse question, children used the informants' expertise on 72% of the trials, with a distribution significantly different from chance, $\chi^2(2, n = 21) = 7.72$, $p = .021$. On the ask question, children only used the informants' expertise on 65% of the trials, with a distribution not significantly different from chance.

Discussion

These data suggest that children understood the expertise of the confederates and that 4-year-olds recognize that individuals have access to different pieces of causal knowledge. These data also suggest that 4-year-olds have some understanding that another's knowledge can be used to make novel inferences. In particular, children can use individuals' reliability about objects' causal properties to make inferences about those objects' labels.

Although the pattern of responses on the ask and endorse questions was not significantly different from each other, the distribution of responses to the ask question was at chance levels, whereas responses to the endorse question were significantly different from chance. We suggest two possible explanations for this difference. First, previous investigations (e.g., Birch et al., 2008; Jaswal & Neely, 2006; Koenig & Harris, 2005) presented children with a consistently accurate informant and a consistently inaccurate informant. Here, children were asked to monitor the accuracy of different kinds of knowledge. It is possible that requiring children to monitor both informants' accuracy taxed their metacognitive abilities, making the ask question slightly more difficult than the endorse question. Consistent with this idea, Pasquini et al. (2007) asked children to monitor the relative accuracy between the two informants. When one informant was always correct or always incorrect, 4-year-olds' responses to the ask and endorse questions were both above chance. But when both informants responded with different degrees of inaccuracy (one was accurate on 75% of the training trials, the other on 25%), 4-year-olds showed a similar pattern of performance on ask and endorse questions to the results presented here.

Second, unlike much previous research on children's understanding of others' reliability, we showed children live confederates instead of asking them to make inferences about people on videos, characters, or puppets (Corriveau et al., 2009, used a similar procedure). Regardless of the child's choice on the ask question, the informants generated their labels in a fixed order. Thus, children sometimes chose to ask one confederate, but the other confederate offered her label first. It is possible that children thought the ask question referred to whom the experimenter would ask first for a label. We changed this aspect of the procedure in Experiment 2.

In summary, these data suggest that 4-year-olds can both extract and use differential expertise about the causal properties of objects. Children can monitor the quality of information that they hear from the confederates in order to predict which of the two has the relevant information about that object's category membership. At issue is how children are monitoring this information. Our hypothesis is that children are not only monitoring the informants' accuracy but are also integrating their representation of those individuals' reliability with their existing knowledge of the relevance of that information. As children develop knowledge about what information is important for category membership, they should use that knowledge to interpret the reliability information they observe. In order to understand that a plumber's ability to label pipe is independent of that person's ability to craft domestic policy, one must recognize that the causal relations involved in unclogging a drain are different from the causal relations involved in balancing the federal budget.

However, an alternative possibility is that in Experiment 1, children might have just associated the two causal properties with the two individual informants. On this view, children endorse the red expert's label for the object that activated the machine red because they have learned to associate that individual with objects that have particular efficacy. This suggests that 4-year-olds might treat reliable answers to any kind of question about the objects as relevant to the inferences they are asked to make. Experiment 2 examines whether 4-year-olds use any kind of expertise as relevant to making inferences about objects' labels or if they limit their use of the informants' expertise to information related to category membership.

Also, what about younger children? Experiment 1 showed that 4-year-olds are capable of monitoring informants' causal knowledge to make inferences about category membership. Experiment 2 examined whether younger children have similar monitoring abilities. Given that Pasquini et al. (2007) found that 3-year-olds tend to be unable to track individuals' reliability when both are stochastic, 3-year-olds might struggle tracking informants' knowledge.

Experiment 2

Three- and 4-year-olds were introduced to confederates who possessed different pieces of knowledge about unfamiliar objects. This knowledge was either relevant or irrelevant to inferences about those objects' category membership. Instead of presenting children with objects that had different causal properties, we relied on another property of objects that has been found related to their category membership: insides. In Experiment 2, half of the children encountered two informants who were experts about the internal properties of unfamiliar objects. In particular, one expert always knew whether an object had red insides but was ignorant about objects with green insides; the second expert knew when objects had green insides but was ignorant about the objects with red insides. The other half of children learned that the informants were knowledgeable about whether those objects had a particular colored sticker on the back (an analogous nonobvious external property). Several investigations have suggested that objects' insides are related to category membership whereas this kind of external property is not (Gelman & Wellman, 1991; Gottfried & Gelman, 2005; Sobel et al., 2007). We examined whether children would only use expertise related to category membership (i.e., insides) to make inferences about whom to ask for information about object labels and whose labels they should endorse.

Previous investigations have suggested that children develop an understanding of the relation between insides and category membership between the ages of 3 and 4 (e.g., Gottfried & Gelman, 2005; Sobel et al., 2007). Other investigations (e.g., Pasquini et al., 2007) found that 3-year-olds struggled to recognize the relative accuracy of two informants, suggesting that they might also struggle at accurately keeping track of the informants' expertise regardless of the relevance of that information. Given these findings, we predicted an interaction between age and expertise, particularly for responses on the endorse questions: Four-year-olds should relate the confederates' knowledge to the objects' labels only when shown that the confederates have expertise about the internal properties of the objects, but not the external properties of the objects. Three-year-olds should not use that knowledge in either condition.

Method

Participants

Forty 3-year-olds (19 girls, *M* = 42.95 months, *SD* = 3.46, range = 36–48 months) and forty 4-year-olds (19 girls, *M* = 55.35 months, *SD* = 3.40, range = 49–60 months) were recruited from a set of preschools and a list of hospital births. Most children were Caucasian and from middle- to upper-middle-class families; however, no formal measure of SES was taken. The participants did represent the diversity of the population from which they were sampled.

Materials

Six wooden blocks, each different in shape and color, were used. Each block had a 1/2 in. (diameter) hole drilled into it, covered by a wooden dowel. Behind this dowel was either a red or green plastic pin (three of each color), which gave the appearance of an internal part (labeled "stuff" in the procedure). Each block also had a 3/4 in. sticker on the back colored red or green; the color of the sticker matched the color of the internal part. None of the blocks was red or green. Children either saw the internal part or the sticker, and they did not know about the other property of the block.

Procedure

As in Experiment 1, the experimenter introduced the child to two confederates. Children were told they would play a game with a set of toys. In the *insides* condition, children were told that some toys were made out of red stuff and some toys were made out of green stuff. In the *outsides* condition, children were told that some toys had red stickers on the back and other toys had green stickers on the back.

Training trials. The structure of the training trials was similar to Experiment 1. Children were told that the "friends" knew about the toys, so each was asked what they thought about each toy (i.e., what kind of stuff it was made of, or what color its sticker was) one at a time. The objects always had their door in place and were held and brought out from under the table such that neither the child nor the confederates could see the objects' insides or the stickers. One confederate was always correct about an object containing red stuff or having a red sticker on the back and said "I don't know" otherwise. The other confederate was always accurate about an object containing green stuff or having a green sticker on the back and said "I don't know" otherwise. Thus, on each trial, children were told what each confederate thought, then were shown either the internal or external property of the object, and then were asked, "Who was right?" Corrective feedback was given if children answered incorrectly. There were four training trials, two in which the red confederate was correct and two in which the green confederate was correct (order was randomly determined). This denotes a difference between this experiment and Experiment 1. Because we intended to work with 3-year-olds in this experiment, we were concerned with the length of the procedure, and thus gave children fewer training trials.

Test trials. Immediately after the fourth training trial, two additional blocks were introduced one at a time. Dependent on condition, the child was shown each block's internal or external property. One block had a red inside or outside; the other had a green inside or outside (order counterbalanced). As in Experiment 1, children were told that the objects had special names. On each trial, children were asked the same *ask* question (i.e., "Who do you want to ask for the special name of this toy?"). After the child responded, both confederates were asked what they thought the special name of the toy was, and both generated novel words (e.g., *dax* vs. *wug*). Children were asked a similar *endorse* question (i.e., "What do you think the special name of this toy is: a *dax* or a *wug*?"). Finally, after both trials, children were asked two *knowledge* questions—who knew about the toys made up of red or green stuff, or with red or green stickers? As in Experiment 1, the person who served as the red or green expert was determined randomly, and the side of the experimenter on which the red expert sat and the specific words the experts used to label the objects were counterbalanced across children.

Results

Children required feedback on ~4% of the training trials, and 66 of the 80 children required no feedback at all, suggesting that they understood the nature of the training. Responses were scored in the same manner as Experiment 1. Children's scores did not differ between the two ask, the two endorse, or the two knowledge questions (all McNemar χ^2 tests = *ns*). These data were combined and are shown in Table 2.

We first examined responses to the ask questions by analyzing the scores on this question via a 2 (age group) × 2 (condition: insides vs. outsides) analysis of variance (ANOVA). No significant effect of condition was found, but this analysis did find a significant effect of age group with 4-year-olds responding more accurately in general than 3-year-olds, $F(1, 76) = 7.81$, $p = .007$, partial $\eta^2 = .09$. Nonparametric analysis confirmed this result: Four-year-olds were more likely to ask the informant

Table 2
Distribution of Responses to the Ask, Endorse, and Knowledge Questions by Age and Condition (Insides, Outsides) in Experiment 2

	Number of correct responses				
	0	1	2	*M*	*SD*
Insides condition					
3-year-olds					
Ask	6	12	2	0.80	0.62
Endorse	8	9	3	0.75	0.72
Knowledge	6	3	11	1.25	0.91
4-year-olds					
Ask	3	6	11	1.40	0.75
Endorse	2	8	10	1.40	0.68
Knowledge	3	1	16	1.65	0.75
Outsides condition					
3-year-olds					
Ask	3	13	4	1.05	0.61
Endorse	4	13	3	0.95	0.61
Knowledge	9	1	10	1.05	1.00
4-year-olds					
Ask	3	8	9	1.30	0.73
Endorse	6	9	5	0.95	0.76
Knowledge	6	1	13	1.35	0.93

with the appropriate expertise across the conditions, Mann–Whitney *U* = 532.00, *z* = −2.81, *p* = .005, *r* = −.31.

A similar ANOVA performed on responses to the endorse questions revealed a slightly different pattern. Again, no significant effect of condition was found, but a main effect of age group was present with 4-year-olds in general responding more accurately than 3-year-olds, *F*(1, 76) = 4.41, *p* = .039, partial η^2 = .06. In addition, a significant interaction between age group and condition was found, *F*(1, 76) = 4.41, *p* = .039, partial η^2 = .06.

Further investigation revealed that 4-year-olds were significantly more accurate on responses to the endorse question than 3-year-olds in the insides condition, *t*(38) = 2.94, *p* = .003 (one-tailed), Cohen's *d* = 0.95, but showed no significant difference in responses to these questions in the outsides condition, *t*(38) = 0.00, *ns*. Four-year-olds were also more accurate in the insides than the outsides condition, *t*(38) = 1.97, *p* = .028 (one-tailed), Cohen's *d* = 0.64, whereas 3-year-olds' responses were not significantly different between these conditions, *t*(38) = 0.35, *ns*. Nonparametric analyses confirmed these results: Four-year-olds were more likely to use the informants' expertise than 3-year-olds in the insides condition, Mann–Whitney *U* = 107.00, *z* = 2.69, *p* = .005 (one-tailed), *r* = .43, but not the outsides condition, Mann–Whitney *U* = 199.00, *z* = 0.03, *ns*. Similarly, 4-year-olds were more accurate in the insides condition than the outsides condition, Mann–Whitney *U* = 135.00, *z* = 1.89, *p* = .040 (one-tailed), *r* = .30, whereas 3-year-olds showed no difference in responses between these conditions, Mann–Whitney *U* = 166.00, *z* = 1.02, *ns*.

We then examined each group's distribution of responses compared with chance. Only 4-year-olds in the insides condition generated distributions of responses different from chance, χ^2(2, *n* = 20) = 9.60, *p* = .008, for the ask questions and χ^2(2, *n* = 20) = 7.20, *p* = .027, for the endorse questions, all other χ^2(2, *n* = 20) values < 4.41, *ns*.

Four-year-olds were more likely to respond accurately to the knowledge question than 3-year-olds. However, this difference was only marginally significant, *t*(78) = 1.74, *p* = .086, Cohen's *d* = 0.39. This suggests that 4-year-olds might have been better able to monitor the two informants' knowledge. To ensure that the older children's above-chance performance in the insides condition was not just due to this group recognizing who had causal expertise, we explored each group's distribution of responses to the knowledge question. Both age groups responded differently from chance across both conditions on these questions, all χ^2(2, *n* = 20) values > 12.29, all *p*-values < .01. Inspection of Table 2 shows that whereas most 4-year-olds answered these questions accurately, 3-year-olds responses were more bimodally distributed. Thus, we were concerned that the younger children did not understand the expertise information available to them in the training.

As a result, we reanalyzed responses from only the children who responded correctly on the two knowledge questions. This analysis was similar to the analyses presented earlier: Four-year-olds in the insides condition showed a pattern of response significantly different from chance on the endorse questions, χ^2(2, *n* = 16) = 6.38, *p* = .041. On the ask questions, these children's responses were only different from chance at a marginally significant level, χ^2(2, *n* = 16) = 5.50, *p* = .064.

Discussion

Experiment 2 found that 4-year-olds could monitor informants' expertise about both internal and external properties of objects, but they used this information judiciously. They only asked for and endorsed the labels generated by the insides experts. Three-year-olds had slightly more difficulty monitoring informants' expertise. Nevertheless, even the 3-year-olds who did so successfully did not make the same inferences as the older children, particularly in the insides condition.

Why did 4-year-olds use the insides experts as sources of knowledge for the objects' labels, but not the outsides experts? In Experiment 1, we suggested that children recognized that there is a link between objects' causal properties and their category membership. Here, we make a similar argument that 4-year-olds recognize that objects' insides are related to their category membership, whereas the nature of a sticker placed on the back of an object is independent of its category.

An alternative possibility is that children do not make a principled inference about the informants' knowledge of the category membership of the objects. Instead, they simply recognize that knowledge about insides indicates that the individuals are more familiar with the objects than knowledge about outsides. Birch and Bloom (2002), for example, suggested that preschoolers use familiarity to make inferences about individuals' proper names. Although this is a possibility, familiarity and knowledge of category membership seem confounded. In both conditions, the confederates are knowledgeable about a nonobvious property of the object that is revealed during the course of the experiment. If children believe that the confederates in the insides condition are more familiar with these objects because of the nature of the property, then they must recognize that the property is relevant to the objects' conceptual structure.

It is important to note that the inferences 3-year-olds made were different from those made by 4-year-olds. There are two possible reasons for this difference. One is that the younger children were unable to register the different expertise between the two informants. Although 3-year-olds struggled with the knowledge question compared to 4-year-olds, this difference was not significant. Even those 3-year-olds who responded with 100% accuracy on the knowledge questions performed at chance on the ask and endorse questions. These responses potentially reflect their not recognizing the relation between objects' internal or external parts and category membership. The 3-year-olds' data do suggest that children are not responding based on recognizing that one informant is more familiar with one set of objects; Birch and Bloom (2002) found that 2-year-olds could make inferences about familiarity. Future research should attempt to tease apart children's knowledge of conceptual structure from the simpler idea that children register that one individual is more familiar with the objects.

General Discussion

Several experiments have found that preschoolers use informants' past reliability when learning conventional or constructed information. We expand on these findings in two ways. First, children appear to use individuals' ability to predict objects' causal efficacy accurately to make subsequent inferences about a new object's label. By age 4, children can transfer their understanding of speakers' reliability about an object's properties to another inference regarding that object. This is consistent with investigations that have demonstrated that children make inferences about object's intended functions based on speakers' reliability about familiar objects' labels (Birch et al., 2008; Koenig & Harris, 2005).

Second, there appears to be a limit to the extent young children will transfer this knowledge. When shown that individuals were knowledgeable about different internal properties of objects, 4-year-olds used that expertise to make inferences about the label of a new object with the same insides. When shown individuals with expertise about a nonobvious external property, 4-year-olds failed to use that information systematically to make the same inference. Younger children did not use either piece of information. These data are consistent with previous research suggesting that children do not relate object's insides with their category membership or causal properties until age 4 (Gottfried & Gelman, 2005; Sobel et al., 2007). The results from Experiment 2 suggest that 4-year-olds use informants' expertise only when it is relevant to the inference they are asked to make.

These data are also consistent with the hypothesis that 4-year-olds are simply better at monitoring multiple sources of information, similar to the way in which 3-year-olds struggle to understand cases where both informants have probabilistic levels of accuracy (i.e., neither informant is always right or always wrong). Pasquini et al. (2007) and other investigations that consider 3-year-olds' inferences given uncertain data (e.g., Corriveau et al., 2009) interpret their findings as 3-year-olds mistrusting informants who have made any error, regardless of the comparative frequency of those errors. Although children never observe an informant err in these experiments, the younger children might be more likely to treat the experts' ignorance as an incorrect response, given that there is no a priori reason for one informant to only know about one kind of causal property.

Future research should attempt to determine whether the developmental difference observed in Experiment 2 was a result of 3- and 4-year-olds' emerging information-processing abilities or their ability to understand that the informants' expertise is related to the inference(s) they are asked to make. One possibility would be to introduce 3- and 4-year-olds to pictures of familiar stimuli from two distinct classes (e.g., animals and artifacts) and two confederates. One confederate would accurately label all of the animals and be ignorant about the artifacts' labels; the other would show the reverse knowledge. If both 3- and 4-year-olds clearly distinguish animals from artifacts (e.g., as suggested by Gelman, 2003; Rosengren, Gelman, Kalish, & McCormick, 1991, among others), both age groups should have little difficulty when asked who could provide labels for pictures of novel animals and artifacts. This would suggest that such information-processing difficulties are not a likely explanation. In contrast, if a similar developmental difference to Experiment 2 is found, the development of these information-processing abilities is a more likely explanation.

Another question that emerges from these experiments is the exact expectations children have about the relations between objects' causal properties, insides, and category membership. For example, most studies of the relation between insides and category membership present children with a conflict between the object's insides and perceptual similarity (e.g., Gelman & Wellman, 1991; Sobel et al., 2007). In these cases, 3-year-olds use perceptual similarity as a guide to category membership, whereas 4-year-olds use the object's insides. We do not believe that 3-year-olds have no understanding of the relation between object's insides and category membership, but for this object set (simple artifacts that were similar to the ones used by Sobel et al., 2007), 3-year-olds might have less insight into the relation between insides and labels. An open question is whether 3-year-olds would respond differently if the artifacts were made more complicated so that the relevance of their insides was clearer. If preschoolers are just developing the ability to monitor informants' expertise, then a manipulation like this should not matter. If, instead, they are integrating that ability with their understanding of relevance, this might improve 3-year-olds' use of the informants' expertise.

One conclusion we can draw from these data is that by the time children are 4, they possess some of the cognitive abilities to distinguish reliable informants among different domains of knowledge. This might be the basis of the more metacognitive ''division of cognitive labor'' findings by Keil and colleagues in which children recognize that individuals have access to different pieces of knowledge, related to domains in which they are shown to be experts. This is consistent with the fact that 4-year-olds in both experiments showed at least some differences in the distributions of responses to the ask and endorse questions. Deciding what the object is called (the endorse question) requires the child to choose which of the confederates generated the accurate label. In contrast, deciding who knows about the label (the ask question) potentially requires the child to choose which confederate would generate the accurate label, which might require better access to the process by which the child would make this decision.

To conclude, there is little doubt that young children can rely on other people as sources of causal knowledge. Children clearly learn language (or at least word referents), science, religious concepts, and the nature of extraordinary beings by relying on the information provided by others. However, how this is accomplished is not well understood. Harris and Koenig (2006) point out that ''children might . . . be equipped with heuristics or strategies for assessing the quality or plausibility of the testimony that is made available to them'' (p. 519). The present data suggest that children's causal knowledge is part of the ''heuristics and strategies'' that Harris and Koenig suggest are critical to learning from testimony. By the age of 4, children are not just using the most reliable source of information but rather are using their own causal knowledge to judge whether that reliability is relevant.

References

Birch, S. A., & Bloom, P. (2002). Preschoolers are sensitive to speaker's knowledge when learning proper names. *Child Development*, *73*, 434–444.

Birch, S. A., Vauthier, S. A., & Bloom, P. (2008). Three- and 4-year-olds spontaneously use others' past performance to guide their learning. *Cognition*, *107*, 1018–1034.

Bloom, P. (1996). Intention, history, and artifact concepts. *Cognition*, *60*, 1–29.

Bloom, P. (2000). *How children learn the meaning of words*. Cambridge, MA: MIT Press.

Booth, A. E., & Waxman, S. (2002). Object names and object functions serve as cues to categories for infants. *Developmental Psychology*, *38*, 948–957.

Clément, F., Koenig, M., & Harris, P. L. (2004). The ontogenesis of trust in testimony. *Mind and Language*, *19*, 360–379.

Corriveau, K. H., & Harris, P. L. (2009). Preschoolers continue to trust a more accurate informant 1 week after exposure to accuracy information. *Developmental Science*, *12*, 188–193.

Corriveau, K. H., Harris, P. L., Meins, E., Ferneyhough, C., Arnott, B., Elliot, L., et al. (2009). Young children's trust in their mother's claims: Longitudinal links with attachment security in infancy. *Child Development*, *80*, 750–761.

Corriveau, K. H., Meints, K., & Harris, P. L. (2009). Early tracking of informant accuracy and inaccuracy. *British Journal of Developmental Psychology*, *27*, 331–342.

Danovitch, J. H., & Keil, F. C. (2004). Should you ask a fisherman or a biologist? Developmental shifts in ways of clustering knowledge. *Child Development*, *75*, 918–931.

de Saussure, F. (1966). *Course in general linguistics*. New York: McGraw-Hill.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. New York: Oxford University Press.

Gelman, S. A., & Bloom, P. (2000). Young children are sensitive to how an object was created when deciding what to name it. *Cognition*, *76*, 91–103.

Gelman, S. A., & Gottfried, G. M. (1996). Children's causal explanations for animate and inanimate motion. *Child Development*, *67*, 1970–1987.

Gelman, S. A., & Wellman, H. M. (1991). Insides and essences: Early understanding of the non-obvious. *Cognition*, *38*, 213–244.

Gopnik, A., & Sobel, D. M. (2000). Detecting blickets: How young children use information about causal powers in categorization and induction. *Child Development*, *71*, 1205–1222.

Gottfried, G. M., & Gelman, S. A. (2005). Developing domain-specific causal-explanatory frameworks: The role of insides and immanence. *Cognitive Development*, *20*, 137–158.

Graham, S. A., Kilbreath, C. S., & Welder, A. N. (2004). Thirteen-month-olds rely on shared labels and shape similarity for inductive inferences. *Child Development*, *75*, 409–427.

Harris, P. L., & Koenig, M. A. (2006). Trust in testimony: How children learn about science and religion. *Child Development*, *77*, 505–524.

Harris, P. L., Pasquini, E. S., Duke, S., Asscher, J. J., & Pons, F. (2006). Germs and angels: The role of testimony in young children's ontology. *Developmental Science*, *9*, 76–96.

Jaswal, V. K., & Neely, L. A. (2006). Adults don't always know best: Preschoolers use past reliability over age when learning new words. *Psychological Science*, *17*, 757–758.

Keil, F. C., Stein, C., Webb, L., Billings, V. D., & Rozenblit, L. (2008). Discerning the Division of Cognitive Labor: An emerging understanding of how knowledge is clustered in other minds. *Cognitive Science*, *32*, 259–300.

Kemler Nelson, D. G., Almasy, L., Crowley, K., Duke, N., Gardner, J.A., Kiggins, V., et al. (1995). Principle-based inferences in young children's categorization: Revisiting the impact of function on the naming of artifacts. *Cognitive Development*, *10*, 347–380.

Kemler Nelson, D. G., Frankenfield, A., Morris, C., & Blair, E. (2000). Young children's use of function information to categorize artifacts: Three factors that matter. *Cognition*, *77*, 133–168.

Koenig, M., Clément, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science*, *10*, 694–698.

Koenig, M., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development*, *76*, 1261–1277.

Koenig, M. A., & Harris, P. L. (2008). The basis of epistemic trust: Reliable testimony or reliable sources? *Episteme*, *4*, 264–284.

Lutz, D. J., & Keil, F. C. (2002). Early understanding of the division of cognitive labor. *Child Development*, *73*, 1073–1084.

Nazzi, T., & Gopnik, A. (2000). A shift in children's use of perceptual and causal cues to categorization. *Developmental Science*, *4*, 389–396.

Nurmsoo, E., & Robinson, E. J. (2009). Identifying unreliable informants: Do children excuse past inaccuracy? *Developmental Science*, *12*, 41–47.

Pasquini, E., Corriveau, K. H., Koenig, M., & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. *Developmental Psychology*, *43*, 1216–1226.

Rosengren, K., Gelman, S. A., Kalish, C., & McCormick, M. (1991). As time goes by: Children's early understanding of biological growth. *Child Development*, *62*, 1302–1320.

Sabbagh, M. A., & Baldwin, D. A. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschoolers' theory of mind and semantic development. *Child Development*, *72*, 1054–1070.

Smith, L. B., Jones, S. S., & Landau, B. (1996). Naming in young children: A dumb attentional mechanism? *Cognition*, *60*, 143–171.

Sobel, D. M., Yoachim, C. M., Gopnik, A., Meltzoff, A. N., & Blumenthal, E. J. (2007). The blicket within: Preschoolers' inferences about insides and causes. *Journal of Cognition and Development*, *8*, 159–182.

Child Development, March/April 2010, Volume 81, Number 2, Pages 680–681

Manuscripts Accepted for Publication

Family Socioeconomic Status and Consistent Environmental Stimulation in Early Childhood, Robert Crosnoe, Tama Leventhal, R. J. Wirth, Kim Pierce, and Robert Pianta.

Defining Best Practice in Care and Protection of Children in Crisis-Affected Settings: A Delphi Study, Alastair Ager, Lindsay Stark, Bree Akesson, and Neil Boothby.

Inconsistency With Prior Knowledge Triggers Children's Causal Explanatory Reasoning, Cristine H. Legare, Susan A. Gelman, and Henry M. Wellman.

Children's Tattling: The Reporting of Everyday Norm Violations in Preschool Settings, Gordon Ingram and Jesse Bering.

Conversation, Development and Gender: A Study of Changes in Children's Concepts of Punishment, Patrick Leman and Marina Bjornberg.

Integrating Different Perspectives on Socialization Theory and Research: A Domain-Specific Approach, Joan Grusec and Maayan Davidov.

Children's Sleep and Adjustment Over Time: The Role of Socioeconomic Context, Mona El-Sheikh, Ryan J. Kelly, Joseph A. Buckhalt, and J. Benjamin Hinnant.

Antecedents and Correlates of the Popular-Aggressive Phenomenon in Elementary School, Philip Rodkin and Glenn I. Roisman.

The Effects of the 1999 Turkish Earthquake on Young Children: Analyzing Traumatized Children's Completion of Short Stories, Elif Celebi Oncu and Aysegul Metindogan Wise.

Individual and Peer Group Normative Beliefs About Relational Aggression: Developmental Changes and Prediction of Relational Aggression, Nicole E. Werner and Laura G. Hill.

Gender, Values, and Occupational Interests Among Children, Adolescents, and Adults, Erica Weisgram, Rebecca S. Bigler, and Lynn S. Liben.

Children's Understanding of Social-Cognitive and Social-Communicative Aspects of Discourse Irony, Eva Filippova and Janet Wilde Astington.

The Head Bone's Connected to the Neck Bone: When Do Toddlers Represent Their Own Body Topography? Celia A. Brownell, Sara R. Nichols, Margarita Svetlova, Stephanie Zerwas, and Geetha Ramani.

Connecting Cues: Overlapping Regularities Support Cue Discovery in Infancy, Sarah Sahni, Mark Seidenberg, and Jenny Saffran.

Mexican-Heritage Children's Attention and Learning From Interactions Directed to Others, Katie G. Silva, Maricela Correa-Chávez, and Barbara Rogoff.

The Development of Social Essentialism: The Case of Israeli Children's Inferences About Jews and Arabs, Dana Birnbaum, Inas Deeb, Gili Segall, Adar Ben-Eliyahu, and Gil Diesendruck.

Parent Involvement and Children's Academic and Social Development in Elementary School, Nermeen El Nokali, Heather J. Bachman, and Elizabeth Votruba-Drzal.

Do Effects of Early Child Care Extend to Age 15 Years? Deborah Vandell, Jay Belsky, Margaret Burchinal, Laurence Steinberg, Nathan Vandergrift, and NICHD Early Child Care Research Network.

I Like Me if You Like Me: On the Interpersonal Modulation and Regulation of Preadolescents' State Self-Esteem, Sander Thomaes, Albert Reijntjes, Bram Orobio de Castro, Brad J. Bushman, Astrid Poorthuis, and Michael J. Telch.

An Eye for Detail: An Event-Related Potential Study of the Rapid Processing of Fearful Facial Expressions in Children, Petra Vlamings, Lisa M. Jonkman, and Chantal Kemner.

Experience-Based and On-Line Categorization of Objects in Early Infancy, Marc Bornstein and Clay Mash.

On the Progression and Stability of Adolescent Identity Formation, Wim Meeus, Rens van de Schoot, Loes Keijsers, Seth J. Schwartz, and Susan Branje.

Dogs, Bogs, Labs, and Lads: What Phonemic Generalizations Indicate About the Nature of Children's Early Word-Form Representations, Erik Thiessen and Meagan N. Yee.

Investigating the Development of Core Cognitive Skills in Autism: A 3-Year Prospective Study, Elizabeth Pellicano.

Home Media and Children's Achievement and Behavior, Sandra Hofferth.

The Effects of Maltreatment and Neuroendocrine Regulation on Memory Performance, Dante Cicchetti, Fred A. Rogosch, Mark L. Howe, and Sheree L. Toth.

Interracial and Intraracial Contact, School-Level Diversity, and Change in Racial Identity Status Among African American Adolescents, Tiffany Yip, Eleanor K. Seaton, and Robert M. Sellers.

Child Care and the Development of Behavior Problems Among Economically Disadvantaged Children in Middle Childhood, Elizabeth Votruba-Drzal, Rebekah Coley, Carolina Maldonado Carreno, Christine Li-Grining, and P. Lindsay Chase-Lansdale.

Children's Classroom Engagement and School Readiness Gains in Pre-Kindergarten, Nina C. Chien, Carollee Howes, Margaret Burchinal, Robert Pianta, Sharon Ritchie, Donna Bryant, Richard Clifford, Diane Early, and Oscar Barbarin.

Early Elementary School Adjustment of Maltreated Children in Foster Care: The Roles of Inhibitory Control and Caregiver Involvement, Katherine Pears, Philip A. Fisher, Jacqueline Bruce, Hyoun K. Kim, and Karen Yoerger.

Children's Understanding of Ordinary and Extraordinary Minds, Jonathan Lane, Henry Wellman, and E. Margaret Evans.

What Paradox? Referential Cues Allow for Infant Use of Phonetic Detail in Word Learning, Christopher Fennell and Sandra R. Waxman.

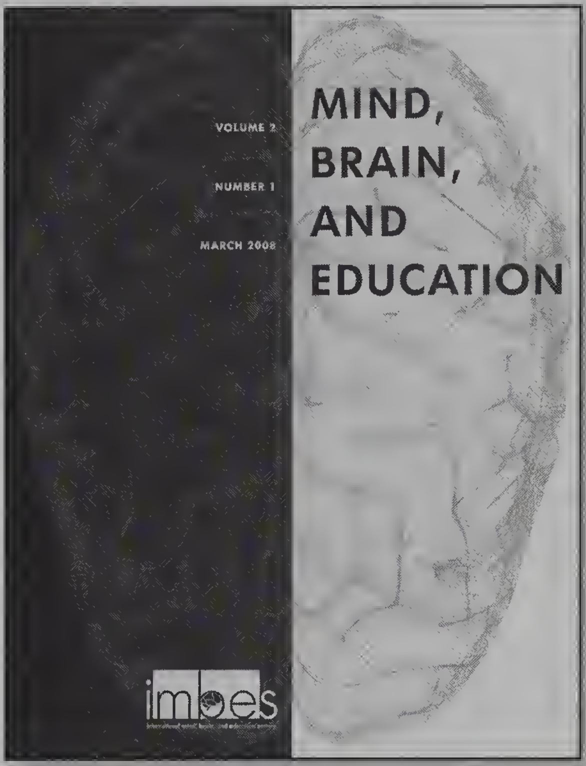
Contagious Yawning in Autistic and Typical Development, Molly Helt, Inge-Marie Eigsti, Peter J. Snyder, and Deborah A. Fein.

Tolerance of Practices by Muslim Actors: An Integrative Social-Developmental Perspective, Maike Gieling, Jochem Thijs, and Maykel Verkuyten.

Structure and Strategies in Children's Educational Television: The Roles of Program Type and Learning Strategies in Children's Learning, Deborah L. Linebarger and Jessica Taylor Piotrowski.

Bridging the Gap: Solving Spatial Means–Ends Relations in a Locomotor Task, Sarah Berger, Karen E. Adolph, and Alisan E. Kavookjian.

Sex Categorization Among Preschool Children: Increasing Utilization of Sexually Dimorphic Cues, Kerri Johnson, Leah Lurye Murphy, and Louis G. Tassinary.



MIND,
BRAIN,
AND
EDUCATION

Edited by
KURT FISCHER *(Editor-in-Chief),*
and **DAVID DANIEL**
(Managing Editor)

The official journal of the International Mind, Brain, and Education Society

Mind, Brain, and Education, recognized as the 2007 **Best New Journal** in the Social Sciences & Humanities by the Association of American Publishers' Professional & Scholarly Publishing Division, encourages submission of papers from the variety of fields relevant to connecting mind, brain, and education in research, theory, and/or practice. The central purpose of the journal is to provide a forum for accessible presentation of basic and applied research on learning and development, including analyses from biology, cognitive science, and education.

WILEY-
BLACKWELL

For more information and to subscribe online visit
www.interscience.wiley.com/journal/mbe

Notice to Contributors

Child Development publishes empirical, theoretical, review, applied, and policy articles reporting research on child development. Published by the interdisciplinary Society for Research in Child Development (SRCD), the journal welcomes relevant submissions from all disciplines. Further information is available at http://www.srcd.org/cd.html.

Types of Articles

Child Development considers manuscripts in formats described below. Inquiries concerning alternative formats should be addressed to the Editor prior to submission. All submissions are expected to be no more than 40 manuscript pages, including tables, references, and figures (but excluding appendices). Authors should provide a justification if the submission is substantially longer. Unless the editor finds that justification compelling, the submission will be returned to the author for shortening prior to editorial review.

Empirical articles comprise the major portion of the journal. To be accepted, empirical articles must be judged as being high in scientific quality, contributing to the empirical base of child development, and having important theoretical, practical, or interdisciplinary implications. Reports of multiple studies, methods, or settings are encouraged, but single-study reports are also considered. Empirical articles will thus vary considerably in length, but should be no longer than 40 manuscript pages; text and graphics should be as concise as material permits. All modes of empirical research are welcome.

Empirical reports are reserved for short cutting-edge empirical papers that are no longer than 4,000 words in length (including text, tables, footnotes, appendices), which advance research and knowledge in an area through noteworthy findings and/or new methods. For manuscripts that require longer descriptions of methods and results, authors should use the Empirical article format.

Reviews focus on past empirical and/or on conceptual and theoretical work. They are expected to synthesize or evaluate a topic or issue relevant to child development, should appeal to a broad audience, and may be followed by a small number of solicited commentaries.

Essays describe original concepts, methods, trends, applications, and theories; these may also be accompanied by solicited commentaries.

Child Development and … are articles that provide readers with tutorials about some new concept or academic specialty pertinent to research in child development. These papers should review the major definitions, methods, and findings of the concept or specialty and discuss past or potential links to child development.

From another perspective is a format in which papers on a focal topic, written by different authors, are published simultaneously. Papers represent diverse perspectives (e.g., authors whose work represents different populations; different disciplines; different theories, methods, or analytic tools). In some cases, calls for submissions on particular topics will be disseminated through SRCD (via e-mail or SRCD publications), and submissions will undergo normal editorial review. In some cases, a submitted manuscript (e.g., an empirical article) may be selected as a lead article for this format, with invited commentaries providing additional perspectives. The editors also welcome suggestions from readers for topics for this format.

Manuscript Submission

Child Development invites for consideration manuscripts that are neither identical to nor substantially similar to work published or under review elsewhere. In the submission cover letter, please provide details about other published or submitted papers having substantial overlap (including data sets) with the new CD submission to enable editors to judge whether the new submission is sufficiently distinct from other work to warrant consideration. Please note if the paper is posted on a website, see http://www. srcd.org/web posting. html. Editors retain the right to reject manuscripts that do not meet established ethical standards for research or dissemination.

Manuscripts should be submitted electronically to the Child Development Online Submission Site at www.srcd.org /CDsubmit/ as a Word or WordPerfect file. Please also submit a cover letter that contains the name(s) of the author(s) and affiliation(s), and the street address, telephone, fax, and electronic mail address of the corresponding author. A corresponding author's submission to *Child Development* implies that all coauthors have agreed to the content and form of the manuscript and that the ethical standards of SRCD have been followed (see the *Child Development* website or pp. 283–284 of the 2000 SRCD Directory). Any financial interest or conflict of interest must be explained to the Editor in the cover letter. The corresponding author is responsible for informing all coauthors, in a timely manner, of manuscript submission, editorial decisions, reviews, and revisions. Please note that if your submission (or its revision) is accepted for publication, each author must submit a signed Exclusive License Form (http://www.srcd.org/Documents/Publications/CD/cdcopyright.pdf) before the manuscript can enter production. Planning ahead to meet this requirement is especially important for manuscripts with multiple authors.

The manuscript file should be formatted with double-spaced, 12-point type, and should include a single paragraph abstract of 100-120 words. Please follow all guidelines on format, style, and ethics provided in the Publication Manual (6th ed.) of the American Psychological Association. Figures included with initial submissions will not be returned. Therefore, please submit only electronic files or copies of figures. If color artwork is submitted, and the authors believe color art is necessary to the presentation of their work, the submissions letter should indicate that one or more authors or their institutions are prepared to pay the substantial costs associated with color art reproduction. Authors should keep a copy of all correspondence, files, and figures to guard against loss.

Manuscript Review

If you have any questions about your submission, please inquire at cdev@srcd.org or call (734) 926-0615. Each manuscript is handled by the Editor or an Associate Editor who consults with one or more Consulting Editors and/or ad hoc reviewers who have relevant expertise. To ensure blind review, cover sheets are removed before review; authors should avoid including any other information about identity or affiliation in submissions. Copies of the submission and associated correspondence are retained in the SRCD archives. For accepted manuscripts, authors are required to prepare a 300-500 layperson's summary for public dissemination purposes. Details are provided to authors as part of final processing.

There is no charge for publication in *Child Development* unless tabular or graphic materials exceed 10% of the total number of pages. Charges are also levied for changes in proofs other than correction of printer's errors. Any inquiries relating to charges or business matters (including reprint orders) should be addressed to Wiley-Blackwell, *Child Development*, Production Coordinator, 350 Main Street, Malden, MA 02148, (781) 388-8200.

Inquiries and suggestions regarding editorial policy may be addressed to: Dr. Jeffrey J. Lockman, Editor, *Child Development*, Department of Psychology, Tulane University, 2007 Percival Stern Hall, New Orleans, LA 70118-5698, lockman@tulane.edu. *Please do not send electronic or paper manuscripts to the above address.* Doing so will delay the processing of your manuscript.

Note to NIH Grantees

Pursuant to NIH mandate, Society through Wiley-Blackwell will post the accepted version of Contributions authored by NIH grantholders to PubMed Central upon acceptance. This accepted version will be made publicly available 12 months after publication. For further information, see www.wiley.com/go/nihmandate.

MONOGRAPHS OF THE SOCIETY FOR RESEARCH IN CHILD DEVELOPMENT

Recent

A DYNAMIC CASCADE MODEL OF THE DEVELOPMENT OF SUBSTANCE-USE ONSET
- *Kenneth A. Dodge, Patrick S. Malone, Jennifer E. Lansford, Shari Miller, Gregory S. Pettit, and John E. Bates* (Serial No. 294, 2009).

FLEXIBILITY IN EARLY VERB USE: EVIDENCE FROM A MULTIPLE-*N* DIARY STUDY
- *Letitia R. Naigles, Erika Hoff, and Donna Vear* (Serial No. 293, 2009).

MARITAL CONFLICT AND CHILDREN'S EXTERNALIZING BEHAVIOR: INTERACTIONS BETWEEN PARASYMPATHETIC AND SYMPATHETIC NERVOUS SYSTEM ACTIVITY
- *Mona El-Sheikh et al.* (Serial No. 292, 2009).

THE EFFECTS OF EARLY SOCIAL-EMOTIONAL AND RELATIONSHIP EXPERIENCE ON THE DEVELOPMENT OF YOUNG ORPHANAGE CHILDREN
- *The St. Petersburg–USA Orphanage Research Team* (Serial No. 291, 2008).

UNDERSTANDING MOTHER-ADOLESCENT CONFLICT DISCUSSIONS: CONCURRENT AND ACROSS-TIME PREDICTION FROM YOUTHS' DISPOSITIONS AND PARENTING
- *Nancy Eisenberg et al.* (Serial No. 290, 2008).

DEVELOPING OBJECT CONCEPTS IN INFANCY: AN ASSOCIATIVE LEARNING PERSPECTIVE
- *David H. Rakison and Gary Lupyan* (Serial No. 289, 2008).

THE GENETIC AND ENVIRONMENTAL ORIGINS OF LEARNING ABILITIES AND DISABILITIES IN THE EARLY SCHOOL YEARS
- *Yulia Kovas, Claire M. A. Haworth, Philip S. Dale, and Robert Plomin* (Serial No. 288, 2007).

THE PRESERVATION OF TWO INFANT TEMPERAMENTS INTO ADOLESCENCE
- *Jerome Kagan, Nancy Snidman, Vali Kahn, and Sara Towsley* (Serial No. 287, 2007).

CHILDREN'S QUESTIONS: A MECHANISM FOR COGNITIVE DEVELOPMENT
- *Michelle M. Chouinard* (Serial No. 286, 2007).

BEST PRACTICES IN QUANTITATIVE METHODS FOR DEVELOPMENTALISTS
- *Kathleen McCartney, Margaret R. Burchinal, and Kristen L. Bub* (Serial No. 285, 2006).

FOUNDATIONS OF SELF-AWARENESS: AN EXPLORATION THROUGH AUTISM
- *R. Peter Hobson, Gayathri Chidambi, Anthony Lee, and Jessica Meyer* (Serial No. 284, 2006).

OUR CHILDREN TOO: A HISTORY OF THE BLACK CAUCUS
OF THE SOCIETY FOR RESEARCH IN CHILD DEVELOPMENT, 1973–1997
- *Diana T. Slaughter-Defoe, Aline M. Garrett, and Algea O. Harrison-Hale* (Serial No. 283, 2006).

PARENTAL SUPPORT, PSYCHOLOGICAL CONTROL, AND BEHAVIORAL CONTROL:
ASSESSING RELEVANCE ACROSS TIME, CULTURE, AND METHOD
- *Brian K. Barber, Heidi E. Stolz, and Joseph A. Olsen* (Serial No. 282, 2005).

BEING HURT AND HURTING OTHERS: CHILDREN'S NARRATIVE ACCOUNTS
AND MORAL JUDGMENTS OF THEIR OWN INTERPERSONAL CONFLICTS
- *Cecilia Wainryb, Beverly A. Brenl, and Sonia Matwin* (Serial No. 281, 2005).

CHILDHOOD SEXUAL ASSAULT VICTIMS: LONG-TERM OUTCOMES AFTER TESTIFYING IN CRIMINAL COURT
- *Jodi A. Quas, Gail S. Goodman, Simona Ghetti, Kristen Weede Alexander, Robin Edelstein, Allison D. Redlich, Ingrid M. Cordon, and David P. H. Jones* (Serial No. 280, 2005).

THE EMERGENCE OF SOCIAL COGNITION IN THREE YOUNG CHIMPANZEES
- *Michael Tomasello and Malinda Carpenter* (Serial No. 279, 2005).

TRAJECTORIES OF PHYSICAL AGGRESSION FROM TODDLERHOOD TO MIDDLE CHILDHOOD:
PREDICTORS, CORRELATES, AND OUTCOMES
- *NICHD Early Child Care Research Network* (Serial No. 278, 2004).

CONSTRAINTS ON CONCEPTUAL DEVELOPMENT: A CASE STUDY OF THE ACQUISITION OF
FOLKBIOLOGICAL AND FOLKSOCIOLOGICAL KNOWLEDGE IN MADAGASCAR
- *Rita Astuti, Gregg E. A. Solomon, and Susan Carey* (Serial No. 277, 2004).

ORIGINS AND EARLY DEVELOPMENT OF HUMAN BODY KNOWLEDGE
- *Virginia Slaughter, Michelle Heron, Linda Jenkins, and Elizabeth Tilse* (Serial No. 276, 2004).

MOTHER CHILD CONVERSATIONS ABOUT GENDER: UNDERSTANDING THE ACQUISITION OF ESSENTIALIST BELIEFS
- *Susan A. Gelman, Marianne G. Taylor, and Simone P. Nguyen* (Serial No. 275, 2004).

WILEY-
BLACKWELL

MAY/JUNE 2010 • VOLUME 81, NUMBER 3

CHILD DEVELOPMENT

Editor
JEFFREY J. LOCKMAN

Associate Editors
MARGARET BURCHINAL • JOHN COLOMBO • MICHAEL CUNNINGHAM • MARY DOZIER
ANDREW J. FULIGNI • MARY GAUVAIN • ROBERTA M. GOLINKOFF • NANCY G. GUERRA
D. GEOFFREY HALL • BETSY HOZA • MELANIE KILLEN • LOUIS J. MOSES • CYNTHIA STIFTER

CURRENT YR/VOL
Marygrove College Library
8425 West McNichols Road
Detroit, MI 48221

SRCD
Society for Research in Child Development

Published by Wiley-Blackwell for the Society for Research in Child Development

Published by Wiley-Blackwell for the Society for Research in Child Development

Child Development

CHILD DEVELOPMENT (ISSN 0009-3920), one of three publications of the Society for Research in Child Development, Inc., is published bimonthly by Wiley Subscription Services, Inc., a Wiley Company, 111 River St., Hoboken, NJ 07030-5774. Periodical Postage Paid at Hoboken, NJ and additional offices.

Postmaster: Send all address changes to *CHILD DEVELOPMENT*, Journal Customer Services, John Wiley & Sons Inc., 350 Main Street, Malden, MA 02148-5020.

Information for Subscribers: *Child Development* is published in six issues per year. Institutional subscription prices for 2010 are: Print & Online: US$670 (US), US$938 (Rest of World), €608 (Europe), £479 (UK). Prices are exclusive of tax. Asia-Pacific GST, Canadian GST and European VAT will be applied at the appropriate rates. For more information on current tax rates, please go to www3.interscience.wiley.com/aboutus/journal_ordering_and_payment.html#Tax. The price includes online access to the current and all online back files to January 1st 1997, where available. For other pricing options, including access information and terms and conditions, please visit www.interscience.wiley.com/journal-info.

Delivery Terms and Legal Title: Prices include delivery of print journals to the recipient's address. Delivery terms are Delivered Duty Unpaid (DDU); the recipient is responsible for paying any import duty or taxes. Legal title passes to the customer on despatch by our distributors.

For more information about Wiley-Blackwell journals, including online access information, terms and conditions, and other pricing options, please visit www.interscience.wiley.com or contact our customer service department, tel: (800) 835-6770 or +1 (781) 388-8598 (Americas); +44 (0) 1865 778315 (Europe, Middle East, and Africa); or +65 6511 8000 (Asia Pacific). Subscription in 2010 includes three issues of *Monographs of the Society for Research in Child Development* and three issues of *Child Development Perspectives.* Single issues from current and recent volumes are available at the current single issue price from csjournals@wiley.com. Earlier issues may be obtained from Periodicals Service Company, 11 Main Street, Germantown, NY 12526, USA. Tel: +1 518 537 4700, Fax: +1 518 537 5899, Email: psc@periodicals.com. Volumes 37 to date are available in microfilm from ProQuest Information and Learning, Serials Customer Service Department, 300 North Zeeb Road, Ann Arbor, MI 48106; telephone: (800) 521-0600 x2873.

Electronic Access: The *Child Development* website can be found at www.srcd.org. Abstract information for this journal is electronically available at www.blackwellpublishing.com/cdev. This journal is available online at Wiley InterScience. Visit www3.interscience.wiley.com to search the articles and register for table of contents e-mail alerts.

Access to this journal is available free online within institutions in the developing world through the HINARI initiative with the WHO. For information, visit www.healthinternetwork.org

Journal Customer Services: For ordering information, claims and any enquiry concerning your journal subscription, please go to interscience.wiley.com/support or contact your nearest office:

Americas: Email: cs-journals@wiley.com; Tel: +1 781 388 8598 or 1 800 835 6770 (Toll free in the USA & Canada).
Europe, Middle East, and Africa: Email: cs-journals@wiley.com; Tel: +44 (0) 1865 778315
Asia Pacific: Email: cs-journals@wiley.com; Tel: +65 6511 8000
Japan: For Japanese speaking support, Email: cs-japan@wiley.com; Tel: +65 6511 8010 or Tel (toll-free): 005 316 50 480.
Further Japanese customer support is also available at http://www.interscience.wiley.com/support
Visit our Online Customer Self-Help available in 6 languages at http://www.interscience.wiley.com/support

Membership communication and address changes should be addressed to the SRCD Executive Office, 2950 South State Street, Suite 401, Ann Arbor, MI 48104, USA. http://www.srcd.org/

Copyright and Photocopying: Journal Compilation © 2010 The Society for Research in Child Development, Inc. All rights reserved. No part of this publication (apart from articles marked "OnlineOpen") may be reproduced, stored or transmitted in any form or by any means without the prior permission in writing from the copyright holder. Authorization to photocopy items for internal and personal use is granted by the copyright holder for libraries and other users registered with their local Reproduction Rights Organization (RRO), e.g., Copyright Clearance Center (CCC), 222 Rosewood Drive, Danvers, MA 01923, USA (www.copyright.com), provided the appropriate fee is paid directly to the RRO. This consent does not extend to other kinds of copying such as copying for general distribution, for advertising and promotional purposes, for creating new collective works or for resale. OnlineOpen articles are made available in accordance with the terms of the Creative Commons Deed, Attribution 2.5 (further details available from www.creativecommons.org), which allows Open Access dissemination of the article, but does not permit commercial exploitation or the creation of derivative works. Special requests should be addressed to PermissionsUK@wiley.com.

Production Editor: Teresa Huang (email: CDEV@wiley.com).

Advertising: Kristin McCarthy (email: kmccarthy@wiley.com).

Indexing/Abstracting: The contents of this journal are indexed or abstracted in the following: *Academic ASAP; Expanded Academic ASAP; Academic Search Elite; Academic Search Premier; Applied Social Sciences Index and Abstracts; Biomedical Reference Collection: Comprehensive Edition; CAB Abstracts; CAB Direct; Contents Pages in Education; Current Contents/Social & Behavioral Science; Current Index to Journals in Education; Cambridge Scientific Abstracts Linguistics & Language Behavior Abstracts; CatchWord; Current Index to Journals in Education; Educational Research Abstracts; Environmental Sciences & Pollution Management; e-psyche; EBSCO Sociological Collection; EBSCO Online; EBSCO Professional Development Collection; Educational Research Abstracts; Educational Resources Information Center; Health and Safety Science Abstracts; ISI Basic Social Sciences Index; InfoTrac OneFile; Ingenta; JCR Social Sciences Edition; Leisure, Recreation, and Tourism Abstracts; Online Computer Library Center FirstSearch Electronic Collections Online; Periodical Abstracts Library; Periodical Abstracts Research II; ProQuest 5000; ProQuest Social Science Plus Text; PsycINFO; Psychological Abstracts; Psychology and Behavioral Sciences Collection; Pollution Abstracts; Safety Science & Risk Abstracts; Sage Family Studies Abstracts; Social Services Abstracts; Social Science Citation Index; Sociological Abstracts; Student Resource Center Bronze; Student Resource Center Gold; Student Resource Center Silver; Violence & Abuse Abstracts; Wilson OmniFile V; Wilson Social Sciences, Abstracts; and Wilson Social Sciences Index*

Published by Wiley-Blackwell for the Society for Research in Child Development

Child Development

Editor

Jeffrey J. Lockman
Tulane University

Associate Editors

Margaret Burchinal, *University of North Carolina at Chapel Hill*
John Colombo, *University of Kansas*
Michael Cunningham, *Tulane University*
Mary Dozier, *University of Delaware*
Andrew J. Fuligni, *University of California, Los Angeles*
Mary Gauvain, *University of California, Riverside*
Roberta M. Golinkoff, *University of Delaware*
Nancy G. Guerra, *University of California, Riverside*
D. Geoffrey Hall, *University of British Columbia*
Betsy Hoza, *University of Vermont*
Melanie Killen, *University of Maryland*
Louis J. Moses, *University of Oregon*
Cynthia Stifter, *The Pennsylvania State University*

Managing Editor

Detra N. Davis

Editorial Assistant

Antonella Caiazza

Consulting Editors

Emma Adam
Karen E. Adolph
Nazan Aksan
William F. Arsenio
Daniel H. Ashmead
Anna Axia
Karen Bartsch
Annie Bernier
Susan Birch
Amy Booth
Heather Bouchey
Paul Boxer
Gregory Stephen Braswell
Mara Brendgen
Gene H. Brody
Keith Burt
Emily W. Bushnell
Gustavo Carlo
Stephanie M. Carlson
Ruth K. Chao
Enrique Chaux
Pablo Chavajay
Tabbye Maria Chavous
Elysia Davis
David Dickinson
Gil Diesendruck
Ruth Feldman
Lisa Gatzke-Kopp
Tamsin C. German
Rick O. Gilmore
Artin Goncu
Susan A. Graham
Elena L. Grigorenko
Laura D. Hanish
Daniel Hart
David B. Henry
Claire H. Hughes
Mutsumi Imai
Deborah J. Johnson
Roger Kobak
Joachim Krueger
Richard Lee
Charlie Lewis
Jeffrey Lidz
Tina Malti
Jelani Mandara
Gary Marcus
Ann S. Masten
Bob McMurray
Julie Meeks-Gardner
Jayanthi Mistry
Ginger A. Moore
Dianna Murray-Close
Velma McBride Murry
Letitia Naigles
Drew Nesdale
Nora S. Newcombe
David Oppenheim
Ana Papafragou
Susan M. Perez
Bradford Pillow
Eva Pomerantz
Douglas Powell
Kimberly K. Powlishta
Samuel Putnam
Suzanne M. Randolph
Robert Reeve
Kristen D. Ritchey
John J. Rieser
James L. Rodriguez
Glenn I. Roisman
Susan A. Rose
Martin D. Ruck
Jenny Saffran
Carlo Schuengel
David Schwartz
Judi G. Smetana
Alan L. Smith
Susan J. Spieker
Timothy Stickle
Dena Phillips Swanson
Adriana Umaña-Taylor
Christopher Trentacosta
Kimberly Updegraff
Marinus H. van IJzendoorn
Brian E. Vaughn
Elisa Velasquez
Lynne Vernon-Feagans
Peter Vishton
Elizabeth Votruba-Drzal
Cecilia Wainryb
Sarah Watamura
Henry M. Wellman
Mike Willoughby
Philip David Zelazo

Child Development, May/June 2010, Volume 81, Number 3, Pages 683–1024

Contents

IN THIS ISSUE 683

ESSAY

Integrating Different Perspectives on Socialization Theory and Research: A Domain-Specific Approach
Joan E. Grusec and Maayan Davidov 687

COMMENTARIES

Commentary and Challenges to Grusec and Davidov's Domain-Specific Approach
Judy Dunn 710

Modules, Theories, or Islands of Expertise? Domain Specificity in Socialization
Susan A. Gelman 715

Domain Specificity in Social Interactions, Social Thought, and Social Development
Elliot Turiel 720

EMPIRICAL REPORT

Connecting Cues: Overlapping Regularities Support Cue Discovery in Infancy
Sarah D. Sahni, Mark S. Seidenberg, and Jenny R. Saffran 727

EMPIRICAL ARTICLES

Do Effects of Early Child Care Extend to Age 15 Years? Results From the NICHD Study of Early Child Care and Youth Development
Deborah Lowe Vandell, Jay Belsky, Margaret Burchinal, Laurence Steinberg, Nathan Vandergrift, and NICHD Early Child Care Research Network 737

The Development of Social Essentialism: The Case of Israeli Children's Inferences About Jews and Arabs
Dana Birnbaum, Inas Deeb, Gili Segall, Adar Ben-Eliyahu, and Gil Diesendruck 757

Gender, Values, and Occupational Interests Among Children, Adolescents, and Adults
Erica S. Weisgram, Rebecca S. Bigler, and Lynn S. Liben 778

The Head Bone's Connected to the Neck Bone: When Do Toddlers Represent Their Own Body Topography?
Celia A. Brownell, Sara R. Nichols, Margarita Svetlova, Stephanie Zerwas, and Geetha Ramani 797

I Like Me If You Like Me: On the Interpersonal Modulation and Regulation of Preadolescents' State Self-Esteem
Sander Thomaes, Albert Reijntjes, Bram Orobio de Castro, Brad J. Bushman, Astrid Poorthuis, and Michael J. Telch 811

Individual and Peer Group Normative Beliefs About Relational Aggression
Nicole E. Werner and Laura G. Hill 826

Antecedents and Correlates of the Popular-Aggressive Phenomenon in Elementary School
Philip C. Rodkin and Glenn I. Roisman 837

The Rise in Cortisol in Family Day Care: Associations With Aspects of Care Quality, Child Behavior, and Child Sex
Megan R. Gunnar, Erin Kryzer, Mark J. Van Ryzin, and Deborah A. Phillips 851

Children's Sleep and Adjustment Over Time: The Role of Socioeconomic Context
Mona El-Sheikh, Ryan J. Kelly, Joseph A. Buckhalt, and J. Benjamin Hinnant 870

Experience-Based and On-Line Categorization of Objects in Early Infancy
Marc H. Bornstein and Clay Mash 884

Mexican-Heritage Children's Attention and Learning From Interactions Directed to Others
Katie G. Silva, Maricela Correa-Chávez, and Barbara Rogoff 898

Children's Understanding of Social-Cognitive and Social-Communicative Aspects of Discourse Irony
Eva Filippova and Janet Wilde Astington 913

Inconsistency With Prior Knowledge Triggers Children's Causal Explanatory Reasoning
Cristine H. Legare, Susan A. Gelman, and Henry M. Wellman 929

Children's Tattling: The Reporting of Everyday Norm Violations in Preschool Settings
Gordon P. D. Ingram and Jesse M. Bering 945

Conversation, Development and Gender: A Study of Changes in Children's Concepts of Punishment
Patrick J. Leman and Marina Björnberg 958

Family Socioeconomic Status and Consistent Environmental Stimulation in Early Childhood
Robert Crosnoe, Tama Leventhal, R. J. Wirth, Kim M. Pierce, Robert C. Pianta, and NICHD Early Child Care Research Network 972

Parent Involvement and Children's Academic and Social Development in Elementary School
Nermeen E. El Nokali, Heather J. Bachman, and Elizabeth Votruba-Drzal 988

The Dynamics of Free Recall and Their Relation to Rehearsal Between 8 and 10 Years of Age
Martin Lehmann and Marcus Hasselhorn 1006

OTHER

Manuscripts Accepted 1021

Errata 1024

Child Development, May/June 2010, Volume 81, Number 3, Pages 683–686

In This Issue

What Is an Effective Parent?

What does it take to be an effective parent? In a theoretical essay, **Grusec and Davidov** (p. 687) bring together various theories of effective parenting by proposing a domain approach to socializing children. In so doing, they suggest that rather than following a single approach to effective parenting, parents operate at different times in different domains (the researchers identify five) that are characterized by different kinds of relationships (for example, protector–protected and teacher–student). Effective parents are those who can figure out in which domain their children are operating and respond in ways that are appropriate to that particular area, with each condition calling for a different form of parenting (comfort, discipline, teaching, etc.). The researchers also note that in each of the domains, parenting is not simply top-down because children also influence their parents' behavior as well as respond to it in a variety of ways. The analysis can help inform research as well as parenting and intervention programs.

Parsing the Environment

Languages are structured and systematic, with recurring and predictable patterns that infants can figure out. **Sahni, Seidenberg, and Saffran** (p. 727) sought to determine whether babies could use one such statistical cue in language to discover an additional, connected cue and thus help them learn language. They conclude that babies are born statisticians, noting that 9-month-olds can use these cues, and in so doing, isolate new cues. Not only can they use these methods to figure out the important patterns in the language, they can use statistical learning to discover the complex structure of language. The study informs our understanding of how very young children acquire language.

Identifying separate entities as belonging to the same group or category is a key skill in developing abstract concepts and acquiring vocabulary. **Bornstein and Mash** (p. 884) ask a number of important questions about the basic processes that very young infants use to categorize, including: Do the methods for assessing category formation in the laboratory reflect the learning that young infants do in more natural environments, such as home? They find that 5-month-olds who saw objects in the laboratory for the first time formed categories of those objects very quickly. However, they also find that natural experience with objects at home is reflected in how children categorize in the laboratory, with infants bringing prior experience to bear in the laboratory setting. By suggesting that infants can learn new categories on the fly, as well as retain and apply categories previously experienced, the findings give us a better understanding of human cognition and development.

Lasting Effects of Child Care

Do children's early child care experiences have any long-term effects on them as adolescents? In the latest installment of the NICHD Study of Early Child Care and Youth Development, **Vandell et al.** (p. 737) sought to determine if early child-care quality, quantity, and type predict children's achievement and behavior problems at age 15. They find that the effects of early child care at age 15 are comparable in size to those previously observed in early childhood and elementary school. Specifically, the longitudinal study of more than 1,300 adolescents finds that teens who attended higher-quality child care scored higher on tests of cognitive and academic achievement than teens who attended lower-quality care. Furthermore, teens who spent more hours in early child care reported more risk taking and greater impulsivity than teens who spent fewer hours in care. The study, the first to document long-term effects of routine, nonrelative care in a large sample of children from economically diverse families, extends the findings of previous research on the topic.

Rodkin and Roisman (p. 837) used longitudinal data on more than 1,000 children from the NICHD study to find that in third to sixth grade, many so-called "tough kids" have an extensive history of early child care before starting school. Tough kids were defined as children who are, perhaps paradoxically, very aggressive and very popular

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8103-0001

with their peers. The study defined extensive child care as averaging 32 hours a week from birth to age 4½, compared to 24 hours a week for others. The findings take into account such factors as gender, ethnicity, mothers' education and sensitivity, and the quality of child care. The study helps illuminate the longer-term effects of extensive child care.

In a study on the quality of child care for children in home-based care, **Gunnar, Kryzer, Van Ryzin, and Phillips** (p. 851) find that child-care providers' behavior is associated with how well children do at child care and how much of the stress hormone cortisol is in their systems. The researchers looked at about 150 mostly White 3- and 4-year-olds in full-time home-based care, comparing their levels of cortisol at home and in child care. Similar to children in center-based care, preschoolers in family day care had elevated levels of cortisol at child care, compared to days when they were at home. Those whose child-care providers were intrusive and over-controlling had larger elevations in cortisol at child care. In girls, this rise was associated with more vigilant and anxious behavior, while for boys it was associated with more angry and aggressive behavior. In contrast, children whose providers were warmer and more supportive showed behavior that was less vigilant, anxious, angry, and aggressive. The findings add to our understanding of how children process stressors in home-based child care, highlighting differences between boys and girls.

Jumping to Conclusions

The belief that all members of a social category (e.g., members of one race) are fundamentally the same, and that you can therefore generalize from one member to all members of a social category is arguably at the foundation of stereotypes, prejudice, and even certain ideologies. **Birnbaum, Deeb, Segall, Ben-Eliyahu, and Diesendruck** (p. 757) carried out two studies of more than 250 kindergarteners, second graders, and sixth graders from three cultural groups in Israel. They find that in Israeli society, this belief is already in place by kindergarten, that its development varies across cultural groups, and that it comes primarily from how language marks the categories rather than from how members of the categories appear visually. By shedding light on the developmental and cultural processes underlying social categorization, the study shows how certain social categories can become powerful sources for children's generalizations about people.

Another study looked at this issue from the perspective of gender. Do men and women tend to work in different fields (think engineering vs. nursing) because they endorse different values? **Weisgram, Bigler, and Liben** (p. 778) examined individuals' occupational values, occupational interests, and perceptions of values associated with different jobs. They studied more than 500 American children (ages 5–10), adolescents (ages 11–17), and young adults (ages 18–23). They found that individuals tended to associate masculine jobs with money, power, and altruism, and feminine jobs with family values. The study also looked at individuals' interest in jobs they encountered for the first time. These ''novel jobs'' were described as being performed by either men or women, and as fulfilling different kinds of values (like wanting to make a lot of money or wanting to have time for one's family). At all ages, interest in novel jobs was influenced by the gender of workers and by the values the jobs were said to fulfill. Overall, the findings suggest that sex differences in occupational values contribute to sex segregation of jobs.

The Physical and Psychological Selves

When do children begin to think about their bodies as they appear to others—as distinct objects, with physical characteristics and parts that are located in specific places? **Brownell, Nichols, Svetlova, Zerwas, and Ramani** (p. 797), based on their study of more than 60 toddlers, find that children are developing a topographic map of their own bodies between 20 and 30 months of age. They define knowledge of body topography as knowing where specific body parts are located, and knowledge about how body parts are arranged or organized in relation to one another. The conclusion—that by age 2½, children have a beginning knowledge of their own bodies' size, shape, and structure—shows us that children's understanding of the self continues to develop in an increasingly refined way in the preschool years.

Past research that has looked at interpersonal influences on children's self-esteem in the long term has found that children's self-esteem is dependent on how much they are liked by others. **Thomaes et al.** (p. 811), looking at interpersonal influences in the short term, find that children's self-esteem can wax and wane as a result of how much they are liked by others in the present. The finding is based on a study of more than 330 eight- to twelve-year-olds in the Netherlands. It suggests that self-esteem is a so-called sociometer, a psychological gauge of how much you are currently valued and accepted by others, and that it is highly dynamic and reacts to current social context, not a stable characteristic as had been previously thought. The study can help inform support programs for socially vulnerable youths.

Relational Aggression

We know that aggression in children causes problems for both those who are aggressive and those who are victimized. In a longitudinal study, **Werner and Hill** (p. 826) examined relational aggression, in which children use relationships to cause harm by, for example, excluding others or spreading rumors. Based on research on more than 240 third through eighth graders, they find that what children think about this type of aggression and what their peers think can influence behavior. Specifically, children's approval of relational aggression grew from elementary to middle school, children who were more approving of this type of behavior engaged in more of it the following year, and students in peer groups that supported the behavior became increasingly aggressive over time. Many schools have programs on physical aggression, but children and adults tend to view relational aggression as more acceptable than physical aggression; these findings provide information for designing school programs.

Counting Sheep, Losing Sleep

Researchers are increasingly interested in the ties between children's sleep and their cognitive and emotional development. **El-Sheikh, Kelly, Buckhalt, and Hinnant** (p. 870) studied how sleep disruptions (namely, the amount, quality, and schedule of sleep) affect children's adjustment. After examining more than 140 children in third to fifth grades, they find that children from poorer families and African American children had higher levels of externalizing symptoms (such as aggression and delinquency) and internalizing symptoms (such as depression, anxiety, and low self-esteem) when they slept poorly. Conversely, when these children slept better, their levels of symptoms were similar to those of other children. This research adds to the mounting evidence linking sleep to day-to-day functioning in childhood, and highlights the role of sleep in a wide range of behavioral problems in high-risk children.

Watch Closely

Children learn a lot from interactions around them that are not addressed directly to them. For example, they learn language by hearing others speak. Do children who are included in the work and social activities of their communities have more opportunities to learn in this way? **Silva, Correa-Chávez, and Rogoff** (p. 898) studied 40 sibling pairs between 5 and 11 years old. Mexican-heritage U.S. children whose families likely are familiar with indigenous Mexican practices paid closer attention to activities around them and learned more. In contrast, Mexican-heritage U.S. children from immigrant families with extensive involvement in Western schooling and practices were not as observant and did not learn as much. The study's findings point to the value of enhancing children's opportunities to learn through observation.

Getting at the Meaning

How early do children understand irony? **Filippova and Astington** (p. 913) studied more than 70 English-speaking Canadian children ages 5, 7, and 9, as well as 24 English-speaking adults. They find that overall, 5-year-olds lagged behind older children in understanding irony, although they did understand some aspects of irony in some conditions—for example, they realized that a sarcastic comment actually means the opposite of what is said. The 7- and 9-year-olds, while able to judge how nice, mean, and funny irony is as well as adults can, were well behind adults in figuring out the ironic speaker's intention or motivation. The study sheds light on early communicative development.

We all like to know why things happen, and children often look for explanations by asking questions, while also making up their own explanations. To determine what motivates children to generate explanations, **Legare, Gelman, and Wellman** (p. 929) studied 80 preschoolers. They find that events that do not square with what they know to be true (i.e., that are inconsistent with prior knowledge) and require them to process new information are a powerful trigger for children to construct explanations. In their search for deeper explanations, they describe what has happened by focusing on underlying reasons that might not be obvious. The research highlights the role that explanations may play in the learning process.

Tattling and Talking

Young children frequently "tattle" on other children who have done something they consider wrong. In the first systematic research to focus on tattling in preschool settings, **Ingram and Bering** (p. 945) find that children (they studied 3- and 4-year-olds in Northern Ireland) tattle to the teacher about once a day and are more likely to tell teachers about bad behavior (hitting someone, damaging property, fighting during playtime) than about behavior that is just different. They also find that children tell the truth in their reports, are more likely to tattle when they themselves are the victims, and vary considerably in their motivations for tattling. And they find that the tattlers are rarely ignored by adults, who support them and intervene when necessary. The study helps us understand the role of tattling, which the authors call an important form of social communication for many young children.

To explore the relations among gender, conversation, and children's concepts of punishment, **Leman and Björnberg** (p. 958) studied about 130 British 9-year-olds. The children listened to stories in which a child had done something wrong, were asked to choose a fair punishment and, in some cases, discussed the topic with other children. Eight weeks later, the children decided on the fairness of the punishment again. Both immediately and 8 weeks later, children who discussed the story with peers made more mature judgments about fairness than those who did not. The researchers suggest that peer conversations can promote moral development because they stimulate children to reflect on a topic after the conversation. Moreover, although the content of the conversations was not influenced by gender, conversation dynamics *were* strongly affected, with boys using more negative interruptions than girls, and children changing their behavior based on whether their conversation partner was a girl or a boy. The study sheds light on how and why peer instruction spurs development and learning.

Promoting Development

The more that environments stimulate children cognitively and intellectually when they are young, the higher their achievement in elementary school. That is what **Crosnoe et al.** (p. 972) find, based on a longitudinal analysis of more than 1,300 children followed from birth in the NICHD Study of Early Child Care and Youth Development. The researchers evaluated children's homes and child care/preschool settings when the children were 4½ years old, studied their first grade classrooms, and evaluated reading and math test scores through fifth grade. Results: What matters is not so much whether parents support children's learning at home or children are stimulated by their teachers in preschool or elementary school but instead whether all *three* things occur. Finally, even though children from advantaged families are more likely to experience this convergence of support for learning across the contexts of their lives, low-income children may benefit more from it. The study has implications for policy, suggesting the value of boosting coordination among parents, preschool classrooms, and elementary schools.

Promoting parent involvement has been a major component of efforts to improve school performance. Extending past research that has focused on parent involvement among preschoolers, **El Nokali, Bachman, and Votruba-Drzal** (p. 988) sought to learn how parent involvement affects children's academic, social, and emotional well-being in elementary school. Based on data from the NICHD study, the researchers examined more than 1,300 children from birth to fifth grade. Children whose parents were more involved across elementary school had fewer problem behaviors and better social skills but showed no changes in achievement. In suggesting that parents continue to wield considerable influence on children's development through elementary school, the study has implications for policies to encourage involvement.

Memory and Order

In school and life these days, children are called on to memorize more and more related and unrelated information. Eight- to 10-year-olds do so by rehearsing words, either individually or together as part of a set. In a longitudinal look at about 75 German second graders, **Lehmann and Hasselhorn** (p. 1006) find that as they get older, children increasingly rehearse several words together. This in turn helps them recall the words by leading them to remember a specific recall order that corresponds to the order of the words when they were presented. The authors' in-depth analysis offers insights into the ways youngsters are developing and improving their capabilities to organize and remember information as they enter into middle childhood.

Anne Bridgman

Child Development, May/June 2010, Volume 81, Number 3, Pages 687–709

Integrating Different Perspectives on Socialization Theory and Research: A Domain-Specific Approach

Joan E. Grusec
University of Toronto

Maayan Davidov
The Hebrew University of Jerusalem

There are several different theoretical and research approaches to the study of socialization, characterized by frequently competing basic tenets and apparently contradictory evidence. As a way of integrating approaches and understanding discrepancies, it is proposed that socialization processes be viewed from a domain perspective, with each domain characterized by a particular form of social interaction between the object and agent of socialization and by specific socialization mechanisms and outcomes. It is argued that this approach requires researchers to identify the domain of social interaction they are investigating, to understand that phenotypically similar behaviors may belong to different domains, and to acknowledge that caregivers who are effective in one type of interaction may not be effective in another.

Developmental psychologists have a long history of interest in how children are socialized or assisted to become mature, competent, and contributing members of the social community. Topics studied and theoretical approaches taken have been diverse. Possibly as a reflection of this diversity, interconnections among various approaches are remarkably few, in spite of the general concern with how agents of socialization support children's development in social, emotional, and intellectual areas. Thus, there has been little attention to understanding how diverse approaches are linked and/or overlap in their implications for the developmental process. This article addresses how such an integration might be accomplished and how various theoretical and research threads might be brought together into a unified conceptualization. We do this by elaborating on a domain-specific approach to the analysis of socialization.

The notion of domain specificity is particularly popular in cognitive psychology, where it has become increasingly clear that information processing is not a unitary, domain-general process. Rather, information from different content areas or types of stimuli is seen as being processed by specialized modules, each of which has evolved to accomplish a recurring important task or function (e.g., face recognition, numerical thinking, grammar processing; Duchaine, Cosmides, & Tooby, 2001; Hirschfeld, 1996; Sperber & Hirschfeld, 2004). Each of these processing modules therefore mediates the acquisition and regulation of a different kind of knowledge or skill, and each is governed by specific processes and mechanisms. A domain analysis has been also applied to socialization, with the realization that different caregiver–child interactions mediate the acquisition of different socioemotional skills and tendencies (Bugental, 2000; Bugental & Goodnow, 1998). The domains in this case are different forms of social relationships or social interaction: Children interact with caregivers and other socialization agents in a variety of ways and contexts, and these interactions can be meaningfully partitioned into several distinct types or domains (Bugental & Goodnow, 1998; Maccoby, 1999). The notion that social relationships—and more specifically, parent–child relationships—are modular has important implications for socialization: Each aspect, or domain, of parent–child interaction is seen as fostering socialization in a different way. Thus, there is no general, all-purpose principle or mechanism of socialization, but rather each form of relationship between the object and the agent of socialization serves a different function, involves different rules and mechanisms for effecting behavior change, and facilitates different outcomes. We note that these domains of social interaction differ from the social knowledge domains explored by

---

Correspondence concerning this article should be addressed to Joan E. Grusec, University of Toronto, Toronto, ON M5S 3G3, Canada. Electronic mail may be sent to grusec@psych.utoronto.ca.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8103-0002

Turiel and his colleagues (1983, 1998), an observation to which we shall return below.

Which domains or forms of social interaction can be identified, between people in general and between children and caregivers in particular? Based on his theoretical and ethnographic work, Fiske (1992) identified four basic types of social relations between individuals: communal sharing (where people treat all members of a category as equivalents, sharing resources and benefits), authority ranking (where people treat one another as hierarchically ordered along a social dimension), equality matching (where people exchange benefits and keep track of imbalances), and market pricing (where people interact based on values determined by a market system). Although this framework is informative for a domain-specific analysis of socialization, it focuses on social relations in general rather than on types of interaction between children and agents of socialization more specifically. For the latter we turn to Bugental and her colleagues (Bugental & Goodnow, 1998; see also Bugental, 2000; Bugental & Grusec, 2006). Based on a review of prior theories and research, they identified four kinds of relationships that exist between agents and objects of socialization, each providing a specific context for socialization. These four domains are: attachment relationships (proximity between caregiver and child in times of stress to maintain the child's safety), social identity relationships (shared identity with members of one's in-group and the acquisition of group norms and routines), hierarchical relationships (caregivers' use of their greater control of resources in order to facilitate children's compliance and discourage disobedience), and reciprocal relationships (mutual, reciprocal exchanges of benefits and positive affect between functional equals). Furthermore, Bugental and her colleagues have reviewed extensive evidence indicating the specificity of these domains. Thus, each type of relationship differs in its evolutionary underpinnings, developmental course, rules of social interaction, regulatory mechanisms including the neurohormonal processes that are activated, type of social information that is sought, and nature of the socialization tasks involved. In the current article we elaborate on this work in several ways. First, we offer some refinements of the typology of domains. Second, we illustrate how this domain analysis helps to bring together the seemingly disparate lines of research on socialization. Third, we focus on what effective parenting entails in each domain, thus unpacking the construct of parental responsivity (Bugental, 2000; Goldberg, Grusec, & Jenkins, 1999). Fourth, we discuss which child outcomes are of greatest relevance to each domain and review evidence supporting domain-specific linkages. Finally, we discuss the implications of a domain approach for the study of socialization, highlighting promising avenues for future research.

In our discussion, we focus on the socialization of children, although noting that any novice who becomes part of a new group, regardless of age, must go through a process of socialization. We also focus on socialization in the family and on parents as primary agents of socialization, observing that although socialization also occurs in other contexts, there is a compelling argument that its primary context is the family (Collins, Maccoby, Steinberg, Hetherington, & Bornstein, 2000). Nevertheless, our analysis of socialization is applicable beyond the family and discussion is confined to the family primarily for the sake of simplicity and because of space limitations. This article begins with a very brief overview of the history of socialization research. It does so for two reasons. First, such a survey highlights the need for bringing together a diversity of approaches that have never been well integrated. Second, such a survey reveals the major themes that have emerged thus far in socialization research, and how they can be meaningfully divided into specific domains.

Socialization Research in Historical Perspective

Reinforcement, Secondary Drives, Discipline, and Parenting Styles

Initial empirical work on socialization was carried out by theorists who were concerned with children's internalization of values, that is, their taking over of the standards and behaviors of society as a means of self-regulating or guiding their own actions. Taking their lead from stimulus-response learning theory principles, they argued that acquired drives underlie most of socialization, with caregivers and their actions taking on secondary reward value through their pairing with primary drive reduction. In this way a dependency drive, as well as a drive to be like or to identify with the parent, develops (Sears, 1957). Much of their work focused on discipline (punishment and reasoning) and its role in the control of children's behavior (Sears, Maccoby, & Levin, 1957). Interest in discipline has continued to the present, with ongoing comparisons of the effectiveness of various discipline techniques for conscience and other forms of social development (e.g., Grusec & Goodnow, 1994;

Hoffman, 1970; Lepper, 1983). Also important is the notion that effective discipline involves the minimization of the child's perceptions of external control so as to promote the child's feelings of self-direction, one aspect of the notion of autonomy (Grolnick, Deci, & Ryan, 1997). One exception to the predominant position that stresses the goal of internalization of "self-chosen" values came in work on behavior management (e.g., Patterson, 1982). This work demonstrated that effective control of external reinforcement contingencies by parents was an important aspect of socialization.

Another related approach appeared in work on parenting styles. Again, the focus was on how parents could best control their children in order to achieve positive socialization outcomes. Firm but responsive control was shown to be generally better than strict and harsh control, or than a lax, laissez-faire approach (Baumrind, 1971). Recent work on parenting styles has focused on behavioral control (monitoring of children's behavior in a context of clearly stated rules) and psychological control (manipulation of children's emotions), with behavioral control associated with better socialization outcomes (Barber, 1996).

We note that common to all these lines of work is that they focus on the relationship between the child and the caregiver as an authority figure who seeks to control or correct the child's behavior. We therefore label this the control domain. It is conceptually similar to Bugental and Goodnow's hierarchical power domain and to Fiske's authority ranking social model.

Observational Learning, Intent Participation, Rituals, and Routines

Bandura and Walters (1963), in what can be seen as a precursor to a domains approach, argued that there are different types of learning, governed by different sets of basic principles. They emphasized the role of observational learning, over that based on reinforcement and acquired drives, as the most fundamental form of learning for humans, and one in which a wide variety of social behaviors are acquired (such as aggression, self-regulation, resistance to temptation, delay of gratification, and moral judgment; Bandura, 1977). Taking a social cognitive stance, Bandura postulated attentional, symbolic, and memorial processes as the underlying mechanisms in observational learning. Moreover, and importantly, he and his colleagues demonstrated that observational learning occurred independent of anticipation of rewards: Children cannot be kept from learning what they have seen (Bandura, Grusec, & Menlove, 1966). Thus, the study of socialization moved from a focus on identification seen as the result of an acquired, secondary drive to a focus on more cognitive mechanisms.

The emphasis on observational learning currently reappears in the attention of anthropologists to learning through intent participation. This is a form of socialization that involves children's close observation and listening-in during valued cultural activities and practices, with the expectation of future participation and the knowledge that such learning is essential for group participation (Lave & Wenger, 1991; Rogoff et al., 2007). Observation with intent participation appears to play a central role in cultures where children are incorporated into adult activities in their community (as opposed to being segregated in formal schooling), but can be a powerful means of socialization in a Western industrialized context as well (e.g., Odden & Rochat, 2005; Rogoff et al., 2007). Researchers interested in cultural practices have stressed that by keenly observing and later participating in group practices, rituals, and routines, children develop their social identity—they develop a sense of belonging to their community and cultural group, and the understanding of what it means to be a member of their in-group (Miller & Goodnow, 1995). Much of this appears to occur in the context of children's most immediate group—the family. Indeed, there is long-standing evidence for the important role of family rituals and routines in conveying to children a sense of social identity and belonging ("this is who we are"), as well as a connection to past and future generations and to the larger social group (Fiese et al., 2002; Wolin & Bennett, 1984).

These lines of thinking and research share a common focus on children's observation and participation in group practices. We therefore label this the group participation domain and note that it overlaps considerably with Bugental and Goodnow's (1998) social identity domain (see also Bugental's, 2000, coalitional group domain). It can also be seen as including certain aspects of Fiske's communal sharing social model.

Attachment Theory

Attachment theorists studied early parent–child relationships and their impact on the child's security, development of the ability to cope with distress, and social competence (Ainsworth, Blehar, Waters, & Wall, 1978). The theory emerged from an ethological and evolutionary perspective rather than the social learning and social cognitive perspectives characteristic of early socialization researchers (Bowlby, 1969/1982). Attachment researchers were interested in how children are brought into the social community or socialized, but whereas social learning theorists such as Sears (1957) had seen the wish to be near the parent as a learned dependency drive, the principles of attachment theory led to different predictions about the impact of parental responsiveness. Thus, for example, attachment theorists argued that parents who responded to the crying of infants ultimately would have babies who cried less whereas learning theorists predicted they would cry more. Because of their opposing principles, attempts to integrate attachment and social learning and social cognitive theory have been virtually nonexistent (Grusec & Ungerer, 2003). We conceptualize attachment theory and relevant work as pertaining to the protection domain, that is, to interactions in which children need and seek comfort, protection, and support from caregivers, and caregivers act to protect children from harm and distress. This domain is conceptually similar to Bugental's attachment domain, and our label reflects the view that protection is a central element of attachment. This domain includes components of Fiske's communal sharing model.

Reciprocal Compliance

There was one notable exception to the failure of social learning, social cognitive, and attachment theorists to take note of each other's thinking. Observing that compliant children were more likely to have caregivers who were responsive to their needs and the proposal that this link was mediated by children's secure attachment (Stayton, Hogan, & Ainsworth, 1973), Maccoby and Martin (1983) suggested instead a direct link between the two: They proposed that parental compliance with children's reasonable needs, wishes, and requests elicits the child's compliance with parental requests in turn, reflecting an inborn tendency to reciprocate the actions of others. This was a route to compliance that differed from that of attachment theorists because it did not involve attachment or protection. In addition, it did not involve the assumption that children and parents were always in conflict and that the latter therefore needed to control the former, but rather that children and parents can cooperate willingly with one another and pursue shared goals. A considerable body of research has since explored the role of parents' positive responding in a context of mutuality (see Kochanska, 2002). We view this direction of theory and research as pertaining to another form of social interaction, that in which children and caregivers cooperate as equals, exchanging and reciprocating one another's bids and favors. We label this the reciprocity domain, similar to Bugental and her colleagues. This domain incorporates components of Fiske's equality matching and communal sharing models.

Teaching and Scaffolding

A final approach to understanding socialization comes from work on instruction and teaching, specifically, on how a more experienced individual works with or guides a less experienced one to acquire more advanced knowledge and skills (Vygotsky, 1978). Central to the notion of guided learning is that the teacher must work within the child's zone of proximal development by presenting problems that are just modestly beyond the child's current level of functioning. As a result of these dialogues the child is able to internalize the language used by the more knowledgeable person and thus develop a more advanced understanding of the task at hand. Although Vygotsky believed that guided learning applies to the acquisition of culturally relevant knowledge and skills in general, most of the research in this area has focused on how children are taught to master cognitive problems. Some evidence for the applicability of supported or guided learning to other types of tasks exists, however. As an example, Turner and Berkowitz (2005) have argued that scaffolding is strongly implicated in moral development and character education. As well, in recent years, there has been a great deal of interest in how parents talk to their children about social and emotional issues, with the plausible assumption that these discussions can also have a didactic purpose in facilitating children's more advanced knowledge and ability in these areas (see Laible & Panfile, 2009, for a review). We view teaching interactions as involving unique features and a distinct context for socialization, and label this domain the guided learning domain. This domain does not clearly map onto Fiske's models, nor does it appear in Bugental's typology. Although it may be interpreted as implicitly subsumed under Bugental and Goodnow's social identity domain, we argue that one-on-one teaching interactions are different in important ways from observation and participation in group rituals and routines (see below).

Integration

This overview underlines the fact that approaches to the study of socialization have been fragmented. Moreover, we note that although children's relationships with parents and other socialization agents have always been at the center of socialization research, each line of study has typically focused on a different aspect of the relationship or a particular way of relating to others (hierarchical power, common group membership, provider/receiver of protection, equal partnership, or teacher/student), that is, on a different domain of social relationships. And work in each domain has called on a different set of principles of learning (reinforcement and punishment, observational learning and social identity, innate tendency to form attachments, reciprocity, and guided learning) to explain how socialization occurs. This fragmentation can be resolved, however, by adopting a domain approach to socialization. By viewing socialization not as involving a single, general mechanism but rather as a collection of different processes pertaining to different elements of the parent–child relationship, seemingly inconsistent lines of research and theory can be brought together. By acknowledging the complexity of the parent–child relationship and the notion that each of its components, or domains, provides children with different opportunities to learn about life and to develop competence, disparate pieces of knowledge can be integrated.

Two caveats are in order. First, our proposed typology of domains is not necessarily exhaustive, as other domains might become evident when research on socialization progresses. Second, although we are presenting the five domains in their prototypic, or pure, form, we also believe that many real-life interactions reflect combinations of or overlaps among domains (see a relevant discussion later in this article). This is consistent with Fiske's (1992) notion that the four elementary models of social relations he identified can be combined in different ways to reflect more complex social situations. Outlining the unique and distinctive features of each domain seems an important first step, however, before considering more complex combinations.

If, as we have argued, socialization involves a variety of domains, then this has implications for the way socialization research is conducted. First, if different principles operate in different domains then researchers need to ensure they identify the domains in which agents and objects of socialization are functioning before talking about the nature of effective socialization. Thus, the notion of ''sensitive and responsive parenting'' requires more precise definition because what constitutes sensitivity and responsivity in one domain may not constitute it in another. Similarly, it is not enough to speak of ''positive socioemotional development'' or ''behavioral adjustment'' because these global terms encompass various aspects relevant to different domains. As well, researchers need to be aware that child outcomes that have surface similarity may belong to different domains, a fact precluding the automatic assumption that phenotypically similar events can be treated as equivalent. Also, researchers (and others) need to be conscious of the fact that agents of socialization who function well in one domain may not necessarily function well in others: While there may be ''good'' parents and ''bad'' parents, one would expect to find many parents who are good in some domains and less good in others. We will elaborate on each of these proposals later in the article.

General Features of Socialization

Before describing features of the five domains relevant to socialization and evidence for the specificity of processes, we make some general observations about the socialization process and the domain approach. As well, we briefly elaborate on differences between the domains of interaction approach discussed in the present article and another influential domain approach in the area of social development, that of domains of social knowledge (Turiel, 1983, 1998).

Socialization, as noted above, is the process by which children acquire the social, emotional, and cognitive skills needed to function in the social community. They acquire these skills through their social interactions for which they have been biologically prepared (Bugental, 2000). As noted above, interactions in each of the five domains involve distinct socialization processes and are pertinent to different child outcomes, which we describe in more detail in the next part of this article. Importantly, socialization is a bidirectional process (Kuczynski, 2003), and this is true in each of the five domains as well as in a more general sense. First, in each domain of social interaction, children have a significant impact on what parents do and on the nature of the interaction. For example, child characteristics influence whether or not parents can easily sooth and comfort the child (protection domain), whether the child makes many requests and bids for attention (reciprocity), how frequently the child's behavior needs to be controlled and which method is most effective (control), what cultural practices and rituals the child is most drawn to (group participation), or how the child needs to be taught in order to master a new task and how quickly he or she can learn (guided learning). Socialization is thus a complex process of interaction between child and parent behavior. In every domain different features of the child that are both environmentally and genetically determined (e.g., temperament, age, past experiences) significantly alter parenting behavior and its effects on children (Grusec & Goodnow, 1994; Grusec & Hastings, 2007). Second, in a more general sense, and very importantly from the perspective of the present article, any parenting effort can be effective only when it is appropriately matched to the domain of interaction in which the child is situated. Thus, the child's need and motivational state must be taken into account for effective socialization to occur. For example, if children are punished when the reason for their behavior is that they are in distress or do not know any better, or if children are comforted when they are not distressed, or guided when they are not motivated to learn or already possess the knowledge that the caregiver is trying to impart, then these parenting interventions are likely to be counterproductive because they are not domain appropriate. According to our domain approach, then, children are important and active contributors to their own socialization.

We also take the position that the domains of socialization presented in this article are universal because they have evolved to solve problems experienced in the human environment of evolutionary adaptedness (Bugental, 2000). Nevertheless, cultural differences are also expected in at least two respects. First, there can be variation in the degree to which a given domain of interaction is frequently engaged in or emphasized in different cultures. For example, guided learning is more widespread in cultures that rely heavily on formal schooling, whereas other cultures may utilize group participation processes to a greater extent (observational learning, intent community participation; Odden & Rochat, 2005; Rogoff et al., 2007). Even in the latter cultures, however, guided learning, though less frequently used, appears to be important in facilitating culturally valued skills that are too complex to master independently (Gauvain, 2005). Second, culture can influence the nature of successful rearing in a given domain of interaction: Certain practices that are successful in one cultural context may appear to be less so in another culture, probably because of different meanings attached to those practices or different goals of socialization emphasized by each cultural group. High levels of rigid control, for example, seem to be more detrimental in Western cultures than in Asian cultures because, in the former, they tend to be associated with parental rejection and lack of warmth, whereas in the latter, they reflect parental care and a desire to instill the important values of hard work, self-discipline, and obedience (Chao, 2002). These differences in emphasis, frequency of use, and in the meanings and consequences of parenting actions, however, do not detract from the basic notion that there are different routes to socialization operative in different domains of interaction and that a parenting practice as interpreted by the child must be appropriate to the domain in which the child is functioning.

Finally, we note that there are both differences and connections between our domains of interaction analysis and Turiel and his colleagues' domains of social knowledge analysis. With respect to differences, the latter line of work focuses on different types of social knowledge (e.g., differentiation among moral, conventional, prudential, and personal issues) and the different forms of reasoning associated with each. The emphasis is on cognition and on how children through their various life experiences acquire a differentiated understanding of different kinds of social issues. The focus of the present article is on different types of interaction with agents of socialization and the different pathways to socialization supported by each of those distinct forms of interaction. The emphasis is on the values, motivations, and skills that children acquire through their social interactions, which then guide their behavior and facilitate their ability to adapt to and function well in their society. Nevertheless there are connections between the two approaches. Most importantly, the fact that children reason differently about different classes of social activities alerts socialization researchers to the fact that not all behaviors are equivalent and that some may be more difficult to modify than others or require different approaches to modification. The personal and prudential domains, for example, are believed by children to involve those activities where authority figures have less right to intervene or interfere than are the moral and social conventional domains (Smetana, 1997). Given that activities in these former domains are considered to be under personal jurisdiction, therefore, it stands to reason that they should be difficult to modify, particularly in the control domain where the parent acts as an authority figure. In the guided learning domain parents act as teachers, and to be effective in this capacity they must understand the child's construal or perception of the situation under discussion. There is evidence, for example, that reasoning or arguments that match the child's way of thinking about a particular class of behavior are more effective than those that do not: Reasoning that refers to harm to others is evaluated more positively in the case of moral misdeeds and reasoning that refers to the maintenance of social order more positively in the case of social conventional misdeeds (see Grusec & Goodnow, 1994; Smetana, 1997, for discussions of this link). Thus, despite their different foci, the two domain approaches do complement one another.

Domains of Socialization: Features and Child Outcomes

We turn now to a description of the five domains relevant to socialization, including discussion of the conditions under which each domain is activated, the nature of the relationship involved in each domain, the different parenting behaviors required as a function of the domain, the child outcomes associated with each domain, and the mechanisms accounting for linkages between parent behavior and child outcome. The order of presentation is loosely guided by developmental considerations. We have chosen to present the protection domain first given that the child's need for protection and the caregiver's provision of protection occurs from the very beginning of life. Mutuality/reciprocity comes next, also because of its appearance early in the life of infants (Adamson & Bakeman, 1991). Control follows as it becomes more necessary once children increase in their mobility and desire for mastery of the environment during the 2nd year of life. Guided learning interactions likely become more prominent toward the end of the 2nd year with the development of language and the accompanying ability to understand and internalize complex explanations (Vygotsky, 1978; see also Thompson, Laible, & Ontai, 2003). Finally, we consider group participation, noting that adherence to the rules of group life and the desire for ritual becomes particularly strong in the 2nd to 4th years of life (Emde, Biringen, Clyman, & Oppenheim, 1991).

Protection

Features of the domain. Many species have evolved so that the young maintain proximity to their caregivers, particularly when they are in danger and need protection. In turn, caregivers have evolved to provide this protection to their offspring, an arrangement that increases the likelihood that parental genes are passed into the gene pool. The protection interaction domain is activated by cues that the child is under real or potential threat, leading the child to seek, and the parent to give, help and support. When caregivers respond sensitively to children's needs for protection, children develop a sense of security and can rely on the caregiver in times of need, whereas insensitive responses lead to lack of confidence in the availability of parental support, and thus to either excessive minimization or augmentation of distress signals (Ainsworth et al., 1978). In the protection domain, then, appropriate parenting involves responding in a way the child perceives to be comforting or helpful when distressed. Perception is important because what is comforting or helpful depends on a variety of features including the child's developmental status and temperament. Other forms of positive parenting such as reciprocity, appropriate discipline, teaching/guided learning, or facilitation of group participation should not contribute to the child's feelings of being protected or secure and able to react appropriately to threat. Similarly, protection and comforting would not be appropriate forms of parenting, as we shall argue subsequently, when the child was not operating in the protection domain. Features of the protection domain are summarized in Table 1 where the parent–child relationship is described as involving provider and recipient of protection, alleviation of the child's distress as the required parent behavior, and confidence in protection as the operative mechanism.

Child outcomes in the domain. An important outcome of appropriate parental protection is that children develop greater ability to respond appropriately to stress and to self-regulate their distress. This ability is at least partially mediated by the normal development of important neurobiological systems, which can be impaired if caregivers are repeatedly insensitive to infants' distress (Gunnar, 2000). In addition, anticipating that support will be available also renders potentially stressful situations less threatening for the child. Indeed, a great deal of research indicates that, compared to their insecurely attached peers, securely attached

Table 1
Domains of Socialization: Parent–Child Relationships, Parental Behavior, and Mechanisms

Domain	Nature of parent–child relationship	Parental behavior that is required	Mechanism of socialization
Protection	Provider–recipient of protection	Alleviate child's distress	Confidence in protection
Reciprocity	Exchange/equality	Comply with child's reasonable requests and influence attempts	Innate tendency to reciprocate
Control	Hierarchical	Use discipline method best suited for achieving parental goal	Acquired self-control
Guided learning	Teacher–student	Match teaching to child's changing level of understanding	Internalization of language and approach used by the teacher
Group participation	Joint members of the same social group	Enable child to observe and take part in appropriate cultural practices	Firm sense of social identity

children, who are assumed to have experienced more sensitive caregiving in the protection domain, are more competent at regulating their distress and negative affect (e.g., Cassidy, 1994). Another outcome is greater empathic capability: Well-protected children are better able to respond empathically to the plight of others because they are able to regulate their own empathic arousal and therefore remain focused on the other's plight rather than on their own personal distress (Davidov & Grusec, 2006a; Eisenberg, Wentzel, & Harris, 1998). This ability to appreciate the emotional plight of others facilitates children's prosocial behavior (Eisenberg & Fabes, 1998) and can also reduce antisocial behavior because the negative consequences of one's actions for others are more easily comprehended (Eisenberg et al., 1999). Finally, the directives and commands of a supportive and reassuring caregiver are more likely to be seen as a sign of caring and, as a result, more likely to lead to compliant behavior (Bretherton, Golby, & Cho, 1997), at least when these directives pertain to children's safety and well-being (see below). It also happens that when caregivers comfort their children they may coincidentally provide models of prosocial and caring practices customary in their culture, and that when they are sensitive to their children's need for protection they may also scaffold their children's acquisition of effective coping strategies (Gottman, Katz, & Hooven, 1996; Thompson, 1994). These latter observations demonstrate how interconnected different domains can be during the socialization process, but do not detract from the message of domain-specificity, because protection is the only domain where making the child feel safe and protected is the key parenting action.

Reciprocity

Features of the domain. This aspect of the parent–child relationship emerges not when the child is distressed but rather when parent and child are interacting as equal-status partners and accommodating one another's wishes as can occur, for example, during play. Reciprocity can take the form of specific exchange of favors (''tit for tat'') as well as nonconditional mutuality, both of which occur in families (Parpal & Maccoby, 1985). From an evolutionary perspective such exchange makes sense because the provision of benefits to kin simultaneously benefits one's own reproductive success, and reciprocal help among nonkin is also often necessary for survival and reproductive success (Hamilton, 1964). In the reciprocity domain, caregivers, by cooperating with children's reasonable demands and bids for attention, set up the conditions whereby children are more likely to cooperate in return. As the parent's willingness to accommodate is reciprocated by the child, the two become attuned and responsive to the needs of each other, sharing common goals that result in a cooperative existence. Child responsiveness in the domain of reciprocity thus reflects a genuine interest in compliance. Indeed and quite importantly, if parents were mistakenly to tangibly reward compliance when it occurs in this domain, external consequences would become salient and the child's motivation to comply would be undermined (Deci & Ryan, 1985). Exchanging or sharing of positive affect also often characterizes these reciprocal interactions (Kochanska, 2002; Parpal & Maccoby, 1985), perhaps because the mutual attunement can be highly pleasing for both interaction partners.

In the reciprocity domain, then, appropriate parenting involves accommodating the child's reasonable requests for compliance. Other forms of positive parenting such as responsiveness to distress, appropriate discipline, guided learning, or encouragement of group participation should not facilitate reciprocated cooperation. Similarly, parental compliance would not be appropriate when the child was not operating in the reciprocity domain. Table 1 describes features of this domain: a parent–child relationship of equality; effective parent behavior, one of compliance with reasonable requests from the child; and a mechanism reflecting the innate tendency of humans to reciprocate favors.

Child outcomes in the domain. Outcomes in this domain involve children's eager willingness to comply with parental requests. Indeed, there is considerable evidence for the importance of parental compliance and mutual reciprocity in socialization. Mothers' willing cooperation with their children's wishes and following of their children's lead during a play interaction predicted children's willing compliance with those mothers during a subsequent clean-up task (Lay, Waters, & Park, 1989; Parpal & Maccoby, 1985). Kochanska and her colleagues (e.g., Kochanska & Aksan, 1995; Kochanska & Murray, 2000) also find that mutually compliant and harmonious relationships during early childhood are predictors of positive social behavior in later childhood. There is also evidence that early cooperative play is associated with fewer conduct problems (Gardner, Ward, Burton, & Wilson, 2003) and that boys whose interactions with their mothers are positively responsive and interconnected have less conflict with their mothers and are less likely to engage in antisocial behavior (Criss, Shaw, & Ingoldsby, 2003). Several proximal mechanisms have been proposed to explain increases in child cooperation and compliance in the presence of parental compliance. One suggestion is that allowing children to control the nature of interactions increases positive mood and thereby encourages positive behaviors (Lay et al., 1989; Parpal & Maccoby, 1985). Alternatively, maternal responsiveness may make the child's identity more salient in an exchange relationship and thereby increase the availability of cooperative scripts in the child's relationship schemas (Parpal & Maccoby, 1985).

In sum, appropriate caregiving in the reciprocity domain results in "receptive compliance." Maccoby and Kochanska and their colleagues have distinguished between situational and receptive compliance, suggesting that the former is more likely to occur when children and parents are in conflict and children do not voluntarily go along with parental requests. In the case of receptive compliance, the child conforms willingly with a parental directive, reflecting a genuine interest in furthering the needs of the dyad.

Control

Features of the domain. From an evolutionary perspective it is important for human social groups to develop definitions of and sanctions for deviance in order to achieve social harmony (Boehm, 2000): To function well within their group individuals must come to accept and obey cultural rules, a requirement that often necessitates the inhibition of personal desires. Children learn these important cultural lessons largely in the control domain. Thus, the control domain involves interactions in which the goals of children and caregivers conflict—children want one thing and caregivers another. This typically happens when children misbehave or are noncompliant as well as during caregiver–child conflict in general. In this domain parents take advantage of those resources over which they have greater control to modify the actions of their children, both to ensure their children's reproductive capability and to encourage socially acceptable behavior. Although good arguments are made for the considerable power that children exert in their relationships with parents (e.g., Kuczynski, 2003), parents do have an arsenal of resources that includes social and material rewards, physical punishment, withdrawal of rewards and privileges, social disapproval, social isolation, and superior knowledge and experience to persuade through reasoning. Thus, they can make use of this power imbalance in various ways to discourage their children's antisocial or undesirable behavior and encourage their prosocial behavior.

There are challenges to be met in this process. First, as noted above, most socialization theorists interested in control have argued that the amount of control being administered must be sufficient to produce the desired behavior but, at the same time, not so forceful or autonomy-threatening that it undermines internalization (Bugental & Grusec, 2006; Deci & Ryan, 1985). This minimization of coerciveness is necessary because an important (although not only) goal of socialization in the control domain is to have children internalize values, that is, to take over parental/societal values as their own so that external pressure or force is no longer necessary: The perception that standards and attitudes guiding behavior have been self-generated is the hallmark of internalization (Hoffman, 1970). An already noted exception to this perspective is Patterson who is less concerned with internalization and who argues that behaviors controlled by external contingencies simply become habitual through repetition (Patterson & Fisher, 2002). In addition, some have argued that internalization of self-chosen values is a less important goal in collectivist cultures, where people's actions are viewed as a reflection of their social roles and obligations rather than their autonomous choices (e.g., Schwartz, 1994). Proper behavior that has been chosen by or for the sake of others, such as behavior motivated by the child's respect for authority or sense of duty toward parents, is therefore highly desirable in these cultures (although a sense of autonomy may still be important in these contexts as well; see Helwig, 2006).

A second challenge in the control domain has to do with the large range of individual differences among children and their misdeeds or antisocial acts. Just as there are differences in what children find distressing and comforting in the protection domain, so too are there differences in what children respond to positively in the control domain. Although earlier work attempted to identify those discipline interventions that work best for all children in all situations, subsequent work has yielded a more complex picture of interactions between specific discipline interventions and features of child and situation. Among the variables shown to moderate the effects of parental control strategies are the child's developmental status, sex, temperament, mood, self-perceptions, and nature of the misdeed committed (Grusec & Goodnow, 1994). As a result, successful parenting in this domain cannot be reduced to the use of a particular ("best") discipline method but rather appears to involve the parent's knowledge of how the particular child views the discipline situation and utilization of that knowledge to select the control strategy that can best achieve the parental goal.

In the control domain, then, appropriate caregiving involves exerting the appropriate amount and manner of authority needed in order to modify the child's behavior in accordance with the socialization agent's goals. If these goals include internalization, this must involve making sure not to overly threaten that child's sense of autonomy; if the goal is habitual compliance, then effective control of reinforcement contingencies is more central; if the goal is deference to authority, then emphasizing the child's social roles and duties is appropriate. Other forms of positive caregiving such as responsiveness to distress, reciprocity, guided learning, or facilitation of group participation should not be very effective at bringing about change in children's behavior and attitudes when children misbehave or are reluctant to cooperate. Similarly, parental discipline, even when mild, would not be appropriate when the child was not operating in the control domain. Table 1 summarizes features of this domain including a hierarchical parent–child relationship, parent discipline that is suitable for achieving the parenting goal, and the acquisition of self-regulation or self-control as the mechanism governing socialization outcomes.

Child outcomes in the domain. Outcomes in the control domain involve children's moral and principled behavior, that is, doing the right thing even when it requires inhibiting one's conflicting goal or desire. Consistent with the discussion above, such behavior can take at least two forms. One is proper conduct that is due to internalization of values. Children engage in principled action because they believe that these actions are self-generated, autonomous, and inherently correct. Thus their behavior is internally motivated and occurs in the absence of surveillance. It is in this case that standards and attitudes are most resistant to change. They can be questioned and they will still be staunchly maintained. Presumably they are subject to modification but not without some effort. A second outcome in the control domain is behavior motivated by more external reasons, such as the child's respect for authority, or the anticipation of rewards or punishments. These can also include "internal" (or introjected) rewards and punishments such as the need to be viewed positively by others or to avoid feeling shame or guilt (Deci & Ryan, 1985; Grolnick et al., 1997). Nor does this second outcome reflect a failure of socialization if it is intended by the agent of socialization. In both cases, however—internally and externally motivated appropriate conduct—socialization depends on the greater control of resources by one member of the dyad than by the other.

Guided Learning

Features of the domain. In the domain of guided learning, socialization agents support children's learning of cognitive, physical, vocational, social, or emotional skills that can improve their ability to adjust and function well in their cultural group. From an evolutionary point of view, social interactions involving teaching and guided learning carry a clear advantage for adaptation. Human beings are less strong physically than many other species and so their superior intellectual skills evolved in order to aid their survival, skills that often require teaching and training. They are also born more immature than most other species and thus they have a protracted period of dependency, which gives caregivers a length of time during which they can pass on information (Gauvain & Perez, 2007). In addition, the human capacity for language enables socialization agents to teach highly complex ideas and skills through verbal communication.

Successful socialization in the guided learning domain requires first that the task being taught is within the child's zone of proximal development; that is, it is a task that the child cannot yet perform independently but could master with the aid of a more experienced other (Vygotsky, 1978). Second, the teacher needs to support, or scaffold (Wood, Bruner, & Ross, 1976), the learning by adjusting his or her guidance to the child's changing skill level and understanding. The teacher's involvement is thus gradually reduced until the child is able to perform the task independently. For this to occur, the teacher needs to assist the child in developing more advanced ways of construing the activity or subject matter at hand; thus, the teacher and student arrive at a shared understanding of the task, or intersubjectivity, and the child internalizes the language, and hence the approach, used by the teacher (e.g., Puntambekar & Hübscher, 2005; Vygotsky, 1978).

In the guided learning domain, then, effective parenting involves the provision of the suitable structure, information, strategies, and feedback that children need in order to learn and improve their skills. Other forms of positive parenting such as responsiveness to distress, reciprocity, appropriate discipline, and encouragement of group participation should not facilitate children's acquisition of complex knowledge and skills that are difficult to learn independently. Similarly, the provision of information within the child's zone of proximal development would not be appropriate when the child was not operating in the guided learning domain. In Table 1 we note that the parent–child relationship in this domain is one of teacher and student, that effective parent behavior requires teaching that is matched to the child's changing level of understanding and mastery of the task (within the zone of proximal development), and that the mechanism of learning is the child's internalization of the language and approach used by the more knowledgeable other.

Child outcomes in the domain. Outcomes in the guided learning domain involve the acquisition of the knowledge or skills being scaffolded (Gauvain, 2005). Moreover and importantly, because sensitive scaffolding allows children to comprehend and become aware of the parent's conception of the issues, strategies being used, and internal processes involved (metacognition and emotion), it has the potential to promote a deep understanding and internalization of the relevant topic or skill (e.g., Gauvain, 2005; Neizel & Stright, 2003; Vygotsky, 1978; see also Thompson et al., 2003).

Although the outcomes assessed in this domain have typically involved cognitive abilities (Gauvain, 2005; Mattanah, Pratt, Cowan, & Cowan, 2005), evidence suggests that guided leaning can also facilitate the acquisition of social and emotional skills. Gottman et al. (1996) have shown that parents' emotional coaching of their children (e.g., teaching their child regarding the nature of emotions, the appropriate expression of emotions, and strategies for managing them), assessed in an interview, predicted children's more adaptive physiological as well as behavioral regulation of emotions. Similarly, maltreating mothers' inability to provide coaching regarding how to cope with emotionally arousing situations contributed to their children's poorer coping with negative affect (Shipman & Zeman, 2001). Studying young infants' interaction skills, Vandell and Wilson (1987) demonstrated that maternal scaffolding of social play was linked to infants' subsequent ability to engage in competent social interaction with siblings and peers. Scaffolding by older siblings also facilitated infants' subsequent social play with peers (see also Leve & Fagot, 1997, for a link between parental scaffolding and peer competence at a later age). As another example, Laible (2004a, 2004b) has shown that mothers who talk to their young children about emotion-laden past events in a coherent and elaborative manner and who scaffold the conversations by providing detail, structure, and feedback, have children who show a more advanced understanding of emotions (see also Fivush, Haden, & Reese, 2006; Laible & Song, 2006; Thompson et al., 2003). Turner and Berkowitz (2005) have argued that many common practices of moral and character education (e.g., "plus-one" moral discussions, class meetings) strongly implicate key elements of scaffolding. Thus, guided learning appears to foster socioemotional as well as cognitive competencies. Another potential outcome of effective scaffolding may involve children's greater confidence regarding their own abilities (Mattanah et al., 2005).

Group Participation

Features of the domain. In the group participation domain, socialization agents can promote children's participation in cultural practices. Humans have an inherent desire to be part of their social group(s), and parents can foster and channel this motivation to facilitate socialization. Evolutionary theorists have suggested that the motivation to be part of the group is located in the need to rely on the larger social group for assistance and for resources (e.g., Brewer, 1999). As the in-group becomes larger and more depersonalized, it becomes necessary to demarcate its boundaries so that benefits or trust are not inadvertently bestowed on the out-group. Thus, symbols and behaviors that differentiate members of the in-group from the out-group become important and are seen not just as different from the out-group's but also as better (Brewer, 1999). Evidence for the importance of group participation to young children comes from children's preference for in-group members and their more positive evaluations of the in-group (Bennett et al., 2004; Dunham, Baron, & Banaji, 2008), their greater perceived similarity to in-group members (Bennett & Sani, 2008), their distress when they are unable to reproduce the behavior of others (Kagan, 1982), and their eagerness to take part in the rituals and routines of their family and classroom (McNamara & Humphry, 2007; Spagnola & Fiese, 2007).

In the group participation domain, children and parents interact as members of a common social group, with parents serving as representative in-group members who convey to children what their group membership, or shared social identity, entails. Thus, parents promote routines, habits, and rituals for the family that reflect and embed many in-group expectations and norms. They also serve as models from whom children learn by observation much about social customs and cultural practices. Importantly, interactions in this domain do not involve the careful scaffolding of children's learning that is characteristic of the guided learning domain. Rather, in this domain children are socialized simply by being exposed to or included as participants in group practices. As noted earlier, the proposal that imitation is a primary form of learning came originally in the work of Bandura and his colleagues (e.g., Bandura, 1977), and subsequent work on learning through intent participation and on rituals and routines has provided an emphasis on group practices (e.g., Fiese et al., 2002; Miller & Goodnow, 1995; Rogoff et al., 2007). Group participation interactions also do not involve the conflict between caregiver and child that is part of the control domain, because these are situations in which children are eager and motivated to participate in their group.

Thus, in the group participation domain, by providing children with opportunities to observe how things are done in the group and by encouraging routines and rituals that reflect and instill these norms, caregivers support children's acquisition of expectations about a myriad of social actions, such as gender-based behavior, how to dress, where to eat and sleep, what is appropriately done in public and what can only be done in private, and so on. Caregivers can further influence this process indirectly by managing the child's environment so as to prevent exposure to undesirable models and routines and encourage exposure to desirable ones. Accordingly, to the best of their ability, parents select the neighborhoods in which their children live and the friends with whom their children interact as well as their access to media presentations. Another important feature of this domain is that children can belong to, and may wish to participate in, different social categories and groups. Caregivers can influence which in-groups are seen as more central or important (e.g., same ethnicity or nationality or religious or political affiliation) as well as which reference group is seen as most pertinent at different points in development or for different aspects of behavior (e.g., same-sex peers, older siblings, adults), with consequences for which practices and norms children will adopt.

If caregivers facilitate children's participation in group-relevant practices and norms—through the routines and rituals they promote, their own modeling, and their encouragement of children's exposure to and identification with other positive models — then children are more likely to adopt desirable social customs without questioning. If caregivers are inconsistent, or promote behaviors that are not socially acceptable, then adoption of group norms and identity will be less successful. Other forms of positive parenting such as responsiveness to distress, reciprocity, appropriate discipline, and guided learning should not facilitate children's automatic adoption of customs and attributes characteristic of the group. Similarly, parental emphasis on ritualized activities or encouragement of similarity to others would not be effective when the child was not operating in the group participation domain. As summarized in Table 1, in this domain the parent–child relationship involves membership in the same social group (with the parent, or another socialization agent, functioning as an exemplary in-group member), appropriate parent behavior is that which enables and encourages the child to watch and take part in desirable cultural practices and customs, and the underlying mechanism for socialization is the child's firm sense of social identity—feeling part of his or her social group(s) and knowing "who we are and what we do."

Child outcomes in the domain. Outcomes of effective parenting in the group participation domain involve conformity to and adoption of practices and routines associated with the group. These attitudes and behaviors are readily acquired because they are viewed simply as "the way things are done" and are strongly entrenched because of frequent repetition. This form of learning—through observation and routine—is powerful because the expectations it produces about appropriate standards of behavior are often unexamined and thus relatively impervious to articulation or change (Bourdieu, 1977). Once questioned, however, they lose their hold and may therefore become more prone to rejection than are attitudes and behaviors acquired in the control and guided learning domains. The latter may be more resistant to change or more strongly internalized because they result from apparent self-choice or from a deep intellectual understanding of the issues involved.

Evidence for Specificity

What evidence is there for domain specificity? The empirical evidence is not systematic in nature, given a general lack of interest by many researchers in the different facets of parenting or in a comparison of their utility for particular outcomes. We have searched for studies that do make a comparison and organized them into the different domains whose existence we have hypothesized.

Protection

Del Carmen, Pederson, Huffman, and Bryan (1993) compared sensitive caregiving in the protection domain (maternal soothing and comforting of infants' distress) at 3 months of age with events that we suggest involve the domains of reciprocity (exchanges of positive affect) and guided learning (interactions involving physical objects). They found that comforting predicted attachment security at the end of the child's 1st year, whereas the other forms of responsive parenting did not. More recently, and utilizing a larger sample, McElwain and Booth-LaForce (2006) found that maternal sensitivity to infant distress at 6 months, but not sensitivity to nondistress, predicted later attachment security. Similarly, Leerkes, Blankson, and O'Brian (2009) found that maternal sensitivity to distress at 6 months, but not sensitivity to nondistress, predicted fewer behavior problems and greater social competence in toddlerhood, and, for highly temperamentally reactive children, better affect regulation. As well, Rodriguez et al. (2005) found that maternal responsiveness during an emotionally demanding or stressful task was predictive of children's ability to self-regulate or delay gratification 3 years later, whereas responsiveness during a task that was not emotionally demanding did not predict self-regulation. Davidov and Grusec (2006a) reported that responsiveness to distress predicted self-regulation of negative affect as well as empathy and prosocial behavior toward others in distress whereas warmth did not. Finally, Hubbard and van IJzendoorn (1991; see also van IJzendoorn & Hubbard, 2000) attempted to explain their failure to replicate Bell and Ainsworth's (1972) findings: Unlike Bell and Ainsworth, they found that infants subsequently cried more, not less, when their mothers responded promptly to their crying. Hubbard and van IJzendoorn hypothesized that prompt response to significant distress cries should lead to less crying in young infants whereas a similar prompt response to instrumental cries or mild discomfort should predict greater amounts of crying. This proposition is certainly consistent with a domain-specific approach if one assumes that significant distress cries are associated with the protection domain and mild or instrumental cries with the control domain. In sum, then, we suggest there is evidence that a range of behaviors that are influenced by maternal responsiveness to distress, that is, by parenting in the protection domain, are not influenced by mutual reciprocity, guided learning, or discipline/control. As well, they are not influenced by responsiveness to nondistress or to warmth, variables that appear to involve more than one domain, as will be discussed below.

Reciprocity

Evidence for specificity in this domain comes from three studies. Parpal and Maccoby (1985) reported that, compared to mothers who played as they normally did, mothers who were trained to minimize directions and questioning during a play period, allowing their children to lead the interaction, had children who subsequently complied more with their requests. In contrast, maternal expressions of warmth in this same situation did not contribute to children's compliance. In a second study, Davidov and Grusec (2006b) showed that children whose mothers were more likely to comply with their requests and bids for attention were in turn more likely to comply willingly with a maternal request to clean up a playroom. Mothers' knowledge regarding their children's reactions to discipline strategies did not contribute to children's willing compliance. Third, Davidov and Grusec (2008) found that mothers' willingness to cooperate with their children's requests, as assessed in an interview, predicted children's receptive compliance with maternal requests, whereas maternal responsiveness to distress did not. In sum, then, we conclude that receptive compliance is influenced by mutual reciprocity but not by responsiveness to distress, discipline, or warmth.

Control

Pettit, Bates, and Dodge (1997) reported that gentle discipline predicted fewer behavior problems and greater social skillfulness (e.g., awareness of the impact of one's actions on others) in children whereas warmth and involvement in children's contacts with peers did not. Davidov and Grusec (2006b) found that mothers who were knowledgeable about the way their children reacted to discipline were better able to induce their children to comply after they had initially refused. On the other hand, mothers who were high in compliance with their children's requests were no more or less able to induce such compliance following initial protest. Many years ago, Hoffman (1970), on the basis of a survey of the research literature, concluded that certain forms of discipline worked well in promoting inhibition of antisocial behavior whereas there was little evidence that observing the restrained behavior of other group members produced such inhibition. Finally, we include the suggestion of Hubbard and van IJzendoorn (1991) that limit setting (e.g., nonresponding) reduces instrumental crying (but not distress cries), whereas a protective, comforting response yields the opposite pattern. In sum, then, we suggest there is evidence that a range of behaviors that are influenced by control and discipline are not influenced by reciprocal compliance, enabling the child to observe appropriate cultural practices, responsiveness to distress, or warmth.

Guided Learning

Laible (2004b) reported that maternal scaffolding (evidenced by an elaborative discourse style) during conversations about emotion-laden topics predicted children's behavioral internalization in a resistance to temptation task, whereas security of attachment (presumably reflective of sensitivity in the protection domain) did not. Similarly, Bornstein and Tamis-LeMonda (1997) found that mothers who responded either physically or verbally in an unobtrusive way when their infant was orienting visually to a person or an object or vocalizing (most likely an example of scaffolding) showed better attention span and symbolic play, whereas responsiveness to distress did not predict attention span or quality of symbolic play. Laible and Song (2006) observed that mothers' elaborative style (scaffolding) when discussing emotion-laden past events with their young children had a significant main effect on children's emotional understanding, whereas shared positive affect during the interaction—a measure primarily reflective of reciprocity—did not. Successful cognitive functioning has been predicted by parental scaffolding but not parental warmth and control (Mattanah et al., 2005) or maternal sensitivity and warmth during play or children's attachment security (Mulvaney, McCartney, Bub, & Marshall, 2006). As well, Bigelow, MacLean, and Proctor (2004) found that mothers' level of scaffolding of infant play was unrelated to their degree of following the infant's lead during play (a behavior relevant to the reciprocity domain), with scaffolding appearing to be more strongly linked to more advanced forms of infant play. In sum, it appears that a variety of outcomes influenced by teaching and guided learning are not influenced by other features of caregiving, including responsiveness to distress, reciprocity, control, and warmth.

Group Participation

Grusec, Goodnow, and Cohen (1997) reported that children who did work around the house on a routine basis (e.g., automatically setting the dinner table every night) showed greater spontaneous concern and consideration for others both in the family and in other settings than did children who did work around the house primarily in response to parental directives and requests. If one assumes that routinely helpful children were operating in the group participation domain while doing household work whereas those who were helpful in response to requests were operating in the control or reciprocity domain, then there is support for the contention that spontaneous concern for others is influenced by group participation and not by reciprocal compliance or discipline. The only other study we have found that enables a comparison between parenting in the group participation and in other domains is by Friedman and Weissbrod (2004), who reported that emerging adults' intentions to initiate family rituals in their future families was significantly correlated with the degree to which their same-sex parent promoted family rituals while they were growing up, but not with other dimensions of parenting such as controlling behavior or caring that reflected responsiveness to distress and warmth. Thus, there is initial evidence that behaviors that are influenced by observation of and participation in group practices are not influenced by other features of caregiving including reciprocity, control, responsiveness to distress, and warmth.

Joint Contributions

As well as studies linking different parenting characteristics to different outcomes, there are also some that find independent contributions of these different parenting characteristics to the apparently same outcome. We would have to argue, of course, that the apparently same outcome in these cases reflects a combination or aggregation of similar but not identical outcomes arrived at by different routes or processes. Accordingly, parental responsiveness to distress and parental warmth made distinct contributions to preschoolers' general competence as rated by teachers (Roberts & Strayer, 1987). An elaborative, scaffolding discourse style during discussions of past emotional events and attachment security both made unique contributions to children's emotional understanding (Laible, 2004b), and elaborative discourse and shared positivity both contributed uniquely to children's positive representations of relationships (Laible & Song, 2006). Finally, research into family meals, an important family routine/ritual, has shown that the frequency of joint family meals predicts fewer health-risking behaviors among adolescents even after controlling for other predictive dimensions of parenting such as parental warmth and support in times of need (Eisenberg, Olson, Neumark-Sztainer, Story, & Bearinger, 2004; Fulkerson et al., 2006).

Summary

We have argued that there are five domains involved in the socialization process: protection, reciprocity, control, guided learning, and group participation. Effective parenting in each of these domains requires different behaviors from parents, and a different type of parental responsiveness and sensitivity. There is considerable initial evidence to support the contention that socialization processes differ in these five domains, although additional research is obviously greatly needed. Finally, we have argued that each domain is linked primarily to different socialization outcomes: effective coping during stressful situations and greater empathy and trust in the protection domain, receptive compliance and cooperation in the reciprocity domain, internalized values and situational compliance or deference to authority in the control domain, acquisition of culturally relevant complex knowledge and skills in the guided learning domain, and a sense of social identity and ready adoption of forms of social custom in the group participation domain.

Implications for Socialization Research and Theory

The notion that socialization occurs in different domains of social interaction that are governed by different psychological processes, rather than one domain governed by one set of processes, has implications for the conduct of socialization research and theorizing. The first is that researchers need to specify the domains to which the parenting and child outcome variables they are studying belong in order to identify what specific parenting actions are related to what specific child outcomes. Such specification is needed to provide appropriate hypotheses for testing. The second implication has to do with the quality of parenting: There is no reason to suppose that parenting in different domains of interaction will be of the same quality. Thus, parents who perform well in one domain involving one form of interaction with their child may not necessarily function well in another domain involving a different kind of relationship or interaction with the same child. We discuss these implications in turn.

Parenting and Child Outcome Variables Need to Be Differentiated According to Domain

Close attention to the domains in which agents of socialization and children are operating is not always evident in current work. We discuss examples of parenting and child variables whose more careful separation might serve to sharpen understanding of socialization processes.

Parenting Variables

Responsiveness and sensitivity, warmth, and harshness are parenting variables that continue to attract considerable attention from socialization researchers (Maccoby, 2007). Each of these is considered in turn with respect to how it relates to the five domains of social interaction.

Responsiveness and sensitivity. Parental responsivity is measured in many ways, with each way not necessarily predictive of the same outcomes. Explicit recognition of this fact appears in recent work by Landry, Smith, and Swank (2006) and by Bornstein, Tamis-LeMonda, Hahn, and Haynes (2008). Landry et al. distinguished among four forms of parental responsiveness to infants (acceptance of the infant's needs, positive affective input, support for the infant's foci of attention, and conversation) and found that the different forms of responsiveness were not intercorrelated and that they predicted different child outcomes. Bornstein et al. distinguished different forms of maternal responsiveness (e.g., affirmations, descriptions, imitations/expansions) to four different types of infant bids, and found that these types of responsiveness were not intercorrelated and followed different developmental trajectories. Although the forms of responsiveness identified in these studies are not all directly relevant to socialization, and the classifications are somewhat different from our domain approach, these findings highlight the fact that "responsiveness" or "sensitivity" can assume a variety of different forms and, therefore, the importance of delineating and unpacking the different aspects of this construct. In the case of socialization, these include: sensitive responding to the child's distress in the protection domain, willing cooperation with the child's reasonable requests and bids for attention and influence in the reciprocity domain, responding to the child's misdemeanors in a manner that takes into account the child's construal of the discipline situation in the control domain, providing information and structure within the zone of proximal development that is suitable to the child's current level of functioning in the guided learning domain, and responsiveness to the child's need to be included in the group and to observe and take part in valuable group activities in the group participation domain.

Warmth. Warmth is frequently operationalized as hugging, kissing, expressing affection, expressing positive affect, and praising (e.g., Robinson, Mandleco, Olsen, & Hart, 1995; Rohner, 2005). From the perspective of this article, then, warmth as it is frequently assessed appears to include a number of features that are relevant to different domains. Reciprocation and mutual sharing of positive affect and affection between caregiver and child are relevant to the reciprocity domain. Contingent social reinforcement, or praise, when delivered in order to convey caregiver approval or direct the child's behavior, belongs to the control domain. Positive affect and affection expressed as part of family celebrations and rituals reflects the group participation domain. In addition, kissing, hugging, and other displays of affection that are aimed at comforting an upset child reflect responsiveness to distress and are thus part of the protection domain. If each of these facets of warmth is associated with a different domain it is clear that the variable needs to be unpacked more carefully.

Harshness. Harsh parenting can involve severe punishment, coldness and rejection, intrusiveness, and nonsympathetic reactions to distress. But severe punishment appears relevant to the control domain whereas rejection and nonsupportive behavior may be relevant to the protection or reciprocity domains, depending on whether the rejection pertains to distress or to bids for attention and compliance. Likewise, intrusiveness may pertain to the control, reciprocity, or protection domain, depending on the manner and context in which it is expressed.

Child Outcome Variables

We discuss four child outcome variables in the context of a need for specificity of operationalization.

Compliance. Compliance is a dependent variable that actually has attracted considerable attention with respect to differentiation of its forms. Internalization, as discussed above, is generally operationalized as compliance with parental values in the absence of surveillance (Grusec, 1997). This form of compliance belongs in the control domain. Maccoby and Kochanska and their colleagues (e.g., Kochanska & Aksan, 1995; Maccoby & Martin, 1983) have distinguished between two kinds of compliance that occur when the socialization agent is present—situational and reciprocal or committed—with the former more obviously occurring in response to external pressure and belonging to the control domain, and the latter part of a cooperative and mutually responsive relationship that belongs in the reciprocity domain. And what of compliance in the protection domain? Attachment theorists suggest that children who are securely attached are more likely to be compliant because they trust their parents to ask for compliance that is in their own best interests (Bretherton et al., 1997). The evidence is mixed, however. For example, secure toddlers are more likely to comply with parental directives when they are in a room with potential safety hazards (e.g., an electric fan; Londerville & Main, 1981) or with maternal suggestions during a challenging problem-solving task (Matas, Arend, & Sroufe, 1978), but there is little suggestion that they are more likely to clean up a playroom or not touch attractive forbidden toys (Matas et al., 1978; Van der Mark, Bakersmans-Kranenburg, & van IJzendoorn, 2002). One testable hypothesis, then, is that children who are securely attached, having parents who are responsive to their distress, are more likely to comply when compliance involves their own safety. Similarly, they might be more willing to accept parental direction when in need of protection or support. Thus compliance as a function of trust in the parent to make benevolent demands may extend only to situations that involve the child's protection from harm or distress.

Externalizing problems. Socialization is frequently assessed with the measurement of behavioral problems such as failure to comply and aggression. One widely used measure of these behavioral problems is the Child Behavior Checklist (Achenbach, 1991). This is a well-regarded instrument that continues to show excellent reliability and validity. Nevertheless, it does not differentiate among behaviors that may be linked with different domains. For example, defiance seems relevant to processes in the control domain, whereas cruelty might be linked to the protection domain through deficits in the development of empathy.

Affective self-regulation. Measures of affective self-regulation often do not differentiate between positive and negative affect, or between types of negative emotion (e.g., Shields & Cicchetti, 1997). Our analysis would suggest that negative emotion such as distress or fear is part of the protection domain, negative emotion such as anger part of the control domain, and positive affect such as enjoyment and pleasure part of the reciprocity or group participation domain.

Prosocial behavior. The term *prosocial behavior* is often used to denote a wide range of desirable or "nice" interpersonal behaviors such as helping, sharing, cooperating, befriending, and showing empathy or sympathy with familiar or unfamiliar peers and adults (e.g., Weir & Duveen, 1981). Moreover, children's own conceptions of what it means to be nice overlap only partly with researchers' common operational definitions of prosocial behavior (Greener & Crick, 1999). Given the multitude of behaviors subsumed under this term, researchers need to pinpoint more narrowly the aspect of prosocial behavior of interest to them and its relevant predictors. For example, empathic responding to another's distress is relevant to the protection domain, friendliness and cooperation with familiar others to the reciprocity domain, and inhibiting one's self-interest for the benefit of others, as in sharing or trying to be fair in games, to the control domain.

Distinguishing Among Parenting Variables and Child Outcomes

It is clear that differentiation of apparently similar behavior into different domains is highly desirable. Certainly, if the parenting variable represents a mixture of domains and the child outcome variable is also an amalgam of different domains, then some relation between the two might be expected. But stronger and more conceptually meaningful relations should emerge when domains are more clearly pinpointed. Coincidentally, the knowledge gained by such a fine-grained approach also would be highly useful for the development of effective parenting interventions, by outlining which parenting behaviors are most relevant to, and which are unlikely to impact, specific child problems.

How might researchers go about distinguishing among child outcomes, many of which appear to be similar in form even if they have been acquired in different domains? One way is to consider why children engage in a specific action by, for example, being aware of features of the situation and context in which the behavior takes place (Mischel & Shoda, 1995). As an example, we use Hubbard and van IJzendoorn's (1991) suggestion that strong and mild distress cries should be distinguished, as they pertain to different domains (protection and control, respectively). Similarly, the researcher should be able to identify prosocial behavior that is routinized and polite from that which stems from genuinely expressed empathic concern or from internalized principles of concern for others. A way of doing this would be to ask children why they had behaved in a given way. Reasons for helping actions stemming from the protection domain might include expressions of empathy for the plight of the recipient, those from the control domain might include statements concerning right and proper behavior, and so on.

Quality of Parenting May Not Be the Same in All Domains

Finally, a domain analysis suggests that parents may be successful in one domain of interaction but not in another. Some, for example, may be very good at assuming the role of protector and responding to their children's distress but less good at assuming the role of disciplinarian and applying appropriate levels of control. Some may excel at interacting as equals and complying with their children's requests but be less good as representative members of the joint social group and at providing them with opportunities to observe and participate in desirable group practices. This differential ability occurs because adaptive parenting in each domain of interaction requires different abilities, sensibilities, and resources from parents. These abilities may be linked to different personality characteristics that, of course, are at least partially genetically mediated as well as influenced by past and present environmental influences. In addition to these unique abilities there may also be characteristics of parents that influence their effectiveness in more than one domain. For example, parents who have difficulty with regulation of their own negative emotion may be less adept at soothing and calming their children's distress as well as less able to deal appropriately with an affect-laden discipline situation. That could explain why parents who have nonautonomous attachment representations, a predictor of insensitivity to their children's distress, also tend to be inflexible and psychologically controlling in discipline situations (Verschueren, Dossche, Marcoen, Mahieu, & Bakermans-Kranenburg, 2006). Similarly, parental intrusiveness can potentially negatively impact parenting in several domains, such as the parent's ability to respond sensitively to distress, to go along with the child's wishes, to provide the appropriate amount of guidance in a teaching interaction, or to enable sufficient autonomy in the control domain. As well, lack of social support or high levels of stress are likely to adversely impact parenting in each of the five domains. As a result, some overlaps in parental competence across domains are to be expected. But alongside such commonalities there is also important intraindividual variability in parental competence as a function of domain, and there are specific processes that are unique to each domain. We suggest that domain-general variables should be differentiated from domain-specific variables if developmentalists are to gain a truly adequate understanding of mechanisms involved in socialization.

Interplay Between Domains

Our analysis should not be taken to imply that processes within each domain are isolated from those in other domains. Such a compartmentalized view is overly simplistic: What occurs in one domain of interaction can certainly be affected by processes in another. We briefly discuss four potential kinds of interplay between domains. One type of interplay has in fact been alluded to throughout this article, and it involves instances in which each interaction partner operates in a different domain; for example, a child seeks comfort and reassurance (protection) but the parent interprets the same situation as a power struggle and responds with discipline (control). As noted earlier, parental efforts are unlikely to be effective in these situations, and may well have detrimental effects, because they are not domain appropriate and thus do not address the child's needs. A second type of interplay is when the interaction combines features of two or more domains for both participants. As noted earlier, Fiske (1992) also acknowledged that various real-life situations constitute more complex combinations of the basic relationship models, or domains. Some familiar examples are when a parent comforts and teaches a child how to self-regulate distress (a combination of protection and guided learning), when a parent combines punishment with informative explanations regarding proper behavior (a combination of control and guided learning), or when a parent enforces prohibitions in order to protect the child from harm (combining features of control and protection). A third form of interplay occurs when the interaction moves in quick succession from one domain to another, as when a reciprocal play interaction suddenly turns into a control situation (due to misbehavior) or when a temper tantrum (control) is soon followed by a need for comfort (protection). In these situations, the preceding interaction provides the situational context for the subsequent interaction in the other domain, with potential carryover or interactive effects. A final type of interplay between domains occurs when experiences and skills acquired in the past in one domain affect current interaction in another domain. Here one domain serves as a broader (rather than immediate) context for another. For example, parents should encounter difficulty when trying to teach (either in the control or guided learning domain) a value or behavior that is contrary to one that has been learned in the group participation domain. Thus inducing or teaching children to have healthy eating habits should be challenging when they have been exposed to disorganized mealtime routines or unhealthy eating practices at home, and convincing children that boys and girls can become anything they want can be difficult when they have frequently observed and participated in cultural rituals and routines emphasizing traditional gender roles. Another example of this type of overlap is the finding that early attachment security moderated the relation between maternal use of minimal power assertion and children's conscience development (Kochanska, Aksan, Knaack, & Rhines, 2004): One possible explanation for this finding is that early effective parenting in the protection domain facilitated children's trust in the mother and their ability to regulate negative emotion, and these capacities enabled them later on to attend to maternal gentle directives and arguments in control interactions.

Clearly, there are many possibilities for a complex interplay between domains. In this article we have focused on outlining the processes within each domain, because our main goal was to offer a framework that integrates the seemingly disconnected foci found in the socialization literature. The ultimate challenge for those interested in socialization, however, would be to gain better understanding not only of the separate pathways within each domain but also of the intricate ways in which these pathways combine and interact with one another.

Concluding Remark

Our suggestions in this article run counter to the thinking of some socialization researchers who see considerable interconnectedness between different aspects of socialization (see, e.g., Forman, Aksan, & Kochanska, 2004). The notion of such interconnectedness is an important one to pursue, but we maintain that a comprehensive research agenda must also include attention to important distinctions between these different aspects. There is indeed a growing body of research, described in this article, that indicates the promise of such an approach.

References

Achenbach, T. M. (1991). *The Child Behavior Checklist for ages 4–18*. Burlington: Department of Psychiatry, University of Vermont.

Adamson, L. B., & Bakeman, R. (1991). The development of shared attention during infancy. In R. Vasta (Ed.), *Annals of child development* (Vol. 8, pp. 1–41). London: Kingsley.

Ainsworth, M. D., Blehar, M., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the Strange Situation*. Hillsdale, NJ: Erlbaum.

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.

Bandura, A., Grusec, J. E., & Menlove, F. L. (1966). Observational learning as a function of symbolization and incentive set. *Child Development*, *37*, 499–506.

Bandura, A., & Walters, R. H. (1963). *Social learning theory and personality development*. New York: Holt, Rinehart, & Winston.

Barber, B. K. (1996). Parental psychological control: Revisiting a neglected construct. *Child Development*, *67*, 3296–3319.

Baumrind, D. (1971). Current patterns of parental authority [Special issue]. *Developmental Psychology Monographs*, *4*(1, Pt. 2).

Bell, S., & Ainsworth, M. D. S. (1972). Infant crying and maternal responsiveness. *Child Development*, *43*, 1171–1190.

Bennett, M., Barrett, M., Karakozov, R., Kipiani, G., Lyons, E., Pavlenko, V., et al. (2004). Young children's evaluations of the ingroup and of outgroups: A multinational study. *Social Development*, *13*, 124–141.

Bennett, M., & Sani, F. (2008). Children's subjective identification with social groups: A self-stereotyping approach. *Developmental Science*, *11*, 69–75.

Bigelow, A. E., MacLean, K., & Proctor, J. (2004). The role of joint attention in the development of infants' play with objects. *Developmental Science*, *7*, 518–526.

Boehm, C. (2000). Conflict and the evolution of social control. *Journal of Consciousness Studies*, *7*, 79–101.

Bornstein, M. H., & Tamis-LeMonda, C. (1997). Maternal responsiveness and infant mental abilities: Specific predictive relations. *Infant Behavior and Development*, *20*, 283–296.

Bornstein, M. H., Tamis-LeMonda, C. S., Hahn, C., & Haynes, O. M. (2008). Maternal responsiveness to young children at three ages: Longitudinal analysis of a multidimensional, modular, and specific parenting construct. *Developmental Psychology*, *44*, 867–874.

Bourdieu, P. (1977). *Outline of a theory of practice* (R. Nice, Trans.). Cambridge, UK: Cambridge University Press.

Bowlby, J. (1982). *Attachment and loss: Vol. 1. Attachment*. New York: Basic Books (Original work published 1969).

Bretherton, I., Golby, B., & Cho, E. (1997). Attachment and the transmission of values. In J. E. Grusec & L. Kuczynski (Eds.), *Parenting and children's internalization of values* (pp. 103–134). New York: Wiley.

Brewer, M. G. (1999). The psychology of prejudice: Ingroup love and outgroup hate? *Journal of Social Issues*, *55*, 429–444.

Bugental, D. B. (2000). Acquisition of the algorithms of social life: A domain-based approach. *Psychological Bulletin*, *26*, 187–209.

Bugental, D. B., & Goodnow, J. G. (1998). Socialization processes. In W. Damon (Series Ed.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (5th ed., pp. 389–462). New York: Wiley.

Bugental, D. B., & Grusec, J. E. (2006). Socialization processes. In W. Damon (Series Ed.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (6th ed., pp. 366–428). New York: Wiley.

Cassidy, J. (1994). Emotion regulation: Influences of attachment relationships. *Monographs of the Society for Research in Child Development, 59*, 228–283.

Chao, R. K. (2002). Parenting of Asians. In M. H. Bornstein (Ed.), *Handbook of parenting: Vol. 4. Social conditions and applied parenting* (2nd ed., pp. 59–93). Mahwah, NJ: Erlbaum.

Collins, W. A., Maccoby, E. E., Steinberg, L., Hetherington, E. M., & Bornstein, M. H. (2000). Contemporary research on parenting: The case for nature and nurture. *American Psychologist, 55*, 218–232.

Criss, M. M., Shaw, D. S., & Ingoldsby, E. M. (2003). Mother–son positive synchrony in middle childhood: Relation to antisocial behavior. *Social Development, 12*, 379–400.

Davidov, M., & Grusec, J. E. (2006a). Untangling the links of parental responsiveness to distress and warmth to child outcomes. *Child Development, 77*, 44–58.

Davidov, M., & Grusec, J. E. (2006b). Multiple pathways to compliance: Mothers' willingness to cooperate and knowledge of their children's reactions to discipline. *Journal of Family Psychology, 20*, 705–709.

Davidov, M., & Grusec, J. E. (2008, July). *Different aspects of responsive parenting predict different forms of successful socialization.* Poster presented at the biennial meeting of the International Society for the Study of Behavioral Development, Würzburg, Germany.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior.* New York: Plenum.

Del Carmen, R., Pederson, F. A., Huffman, L. C., & Bryan, Y. E. (1993). Dyadic distress management predicts subsequent security of attachment. *Infant Behavior and Development, 16*, 131–147.

Duchaine, B., Cosmides, L., & Tooby, J. (2001). Evolutionary psychology and the brain. *Current Opinion in Neurobiology, 11*, 225–230.

Dunham, Y., Baron, A. S., & Banaji, M. R. (2008). The development of implicit intergroup cognition. *Trends in Cognitive Sciences, 12*, 248–253.

Eisenberg, N., & Fabes, R. A. (1998). Prosocial development. In W. Damon (Series Ed.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (5th ed., pp. 701–778). New York: Wiley.

Eisenberg, N., Fabes, R. A., Shepard, S. A., Guthrie, I. K., Murphy, B. C., & Reiser, M. (1999). Parental reactions to children's negative emotions: Longitudinal relations to quality of children's social functioning. *Child Development, 70*, 513–534.

Eisenberg, M. E., Olson, R. E., Neumark-Sztainer, D., Story, M., & Bearinger, L. H. (2004). Correlations between family meals and psychosocial well-being among adolescents. *Archives of Pediatrics and Adolescent Medicine, 158*, 792–796.

Eisenberg, N., Wentzel, M., & Harris, J. D. (1998). The role of emotionality and regulation in empathy-related responding. *School Psychology Review, 27*, 506–521.

Emde, R. N., Biringen, Z., Clyman, R. B., & Oppenheim, D. (1991). The moral self of infancy. Affective core and procedural knowledge. *Developmental Review, 11*, 251–270.

Fiese, B. H., Tomcho, T. J., Douglas, M., Josephs, K., Poltrock, S., & Baker, T. (2002). A review of 50 years of research on naturally occurring family routines and rituals: Cause for celebration? *Journal of Family Psychology, 16*, 381–390.

Fiske, A. P. (1992). The four elementary forms of sociality: Framework for a unified theory of social relations. *Psychological Review, 99*, 689–723.

Fivush, R., Haden, C. A., & Reese, E. (2006). Elaborating on elaborations: Role of maternal reminiscing style in cognitive and socioemotional development. *Child Development, 77*, 1568–1588.

Forman, D. R., Aksan, N., & Kochanska, G. (2004). Toddlers' responsive imitation predicts preschool-age conscience. *Psychological Science, 15*, 699–704.

Friedman, S. R., & Weissbrod, C. S. (2004). Attitudes toward the continuation of family rituals among emerging adults. *Sex Roles, 50*, 277–284.

Fulkerson, J. A., Story, M., Mellin, A., Leffert, N., Neumark-Sztainer, D., & French, S. A. (2006). Family dinner meal frequency and adolescent development: Relationships with developmental assets and high-risk behaviors. *Journal of Adolescent Health, 39*, 337–345.

Gardner, F., Ward, S., Burton, J., & Wilson, C. (2003). The role of mother–child joint play in the early development of children's conduct problems: A longitudinal observational study. *Social Development, 12*, 361–378.

Gauvain, M. (2005). Scaffolding in socialization. *New Ideas in Psychology, 23*, 129–139.

Gauvain, M., & Perez, S. M. (2007). The socialization of cognition. In J. E. Grusec & P. D. Hastings (Eds.), *Handbook of socialization* (pp. 588–613). New York: Guilford.

Goldberg, S., Grusec, J. E., & Jenkins, J. (1999). Confidence in protection: Arguments for a narrow definition of attachment. *Journal of Family Psychology, 13*, 475–483.

Gottman, J. M., Katz, L. F., & Hooven, C. (1996). Parental meta-emotion philosophy and the emotional life of families: Theoretical models and preliminary data. *Journal of Family Psychology, 10*, 243–268.

Greener, S., & Crick, N. (1999). Normative beliefs about prosocial behavior in middle childhood: What does it mean to be nice? *Social Development, 8*, 350–363.

Grolnick, W. S., Deci, E. L., & Ryan, R. M. (1997). Internalization within the family: The self-determination theory perspective. In J. E. Grusec & L. Kuczynski (Eds.), *Parenting and children's internalization of values: A handbook of contemporary theory* (pp. 135–161). New York: Wiley.

Grusec, J. E. (1997). A history of research on parenting strategies and children's internalization of values. In J. E. Grusec & L. Kuczynski (Eds.), *Parenting and children's internalization of values: A handbook of contemporary theory* (pp. 3–22). New York: Wiley.

Grusec, J. E., & Goodnow, J. J. (1994). The impact of parental discipline methods on the child's internalization of values: A reconceptualization of current points of view. *Developmental Psychology, 30*, 4–19.

Grusec, J. E., Goodnow, J. J., & Cohen, L. (1997). Household work and the development of children's concern for others. *Developmental Psychology, 32*, 999–1007.

Grusec, J. E., & Hastings, P. D. (Eds.). (2007). *Handbook of socialization*. New York: Guilford.

Grusec, J. E., & Ungerer, J. (2003). Effective socialization as problem-solving and the role of parenting cognitions. In L. Kuczynski (Ed.), *Parent–child relationships as a dyadic process* (pp. 211–228). Thousand Oaks, CA: Sage.

Gunnar, M. R. (2000) Early adversity and the development of stress reactivity and regulation. In C. A. Nelson (Ed.), *Minnesota symposia on child psychology: Vol. 31. The effects of early adversity on neurobehavioral development* (pp. 163–200). Mahwah, NJ: Erlbaum.

Hamilton, W. D. (1964). Genetical evolution of social behavior. *Journal of Theoretical Biology, 7*, 1–52.

Helwig, C. C. (2006). The development of personal autonomy throughout cultures. *Cognitive Development, 21*, 458–473.

Hirschfeld, L. A. (1996). *Race in the making: Cognition, culture, and child's construction of human kinds*. Cambridge, MA: MIT Press.

Hoffman, M. L. (1970). Moral development. In P. H. Mussen (Ed.), *Carmichael's manual of child psychology* (Vol. 2, pp. 261–360). New York: Wiley.

Hubbard, F. O. A., & van IJzendoorn, M. H. (1991). Maternal unresponsiveness and infant crying across the first 9 months: A naturalistic longitudinal study. *Infant Behavior and Development, 14*, 299–312.

Kagan, J. (1982). The emergence of self. *Journal of Child Psychology and Psychiatry, 23*, 363–381.

Kochanska, G. (2002). Mutually responsive orientation between mothers and their young children: A context for the early development of conscience. *Current Directions in Psychological Science, 11*, 191–195.

Kochanska, G., & Aksan, N. (1995). Mother–child mutually positive affect, the quality of child compliance to requests and prohibitions, and maternal control as correlates of early internalization. *Child Development, 66*, 236–254.

Kochanska, G., Aksan, N., Knaack, A., & Rhines, H. M. (2004). Maternal parenting and children's conscience: Early security as moderator. *Child Development, 75*, 1229–1242.

Kochanska, G., & Murray, K. T. (2000). Mother–child mutually responsive orientation and conscience development: From toddler to early school age. *Child Development, 71*, 417–431.

Kuczynski, L. (2003). Beyond bidirectionality: Bilateral conceptual frameworks for understanding dynamics in parent–child relations. In L. Kuczynski (Ed.), *Handbook of dynamics in parent–child relations* (pp. 3–24). Thousand Oaks, CA: Sage.

Laible, D. (2004a). Mother–child discourse surrounding a child's past behavior at 30 months: Links to emotional understanding and early conscience development at 36 months. *Merrill-Palmer Quarterly, 50*, 159–180.

Laible, D. (2004b). Mother–child discourse in two contexts: Links with child temperament, attachment security, and socioemotional competence. *Developmental Psychology, 40*, 979–992.

Laible, D., & Panfile, T. (2009). Mother–child reminiscing in the context of secure attachment relationships: Lessons in understanding and coping with negative emotions. In J. Quas & R. Fivush (Eds.), *Emotion and memory in development: Biological, cognitive, and social considerations. Oxford series in affective science* (pp. 166–195). Oxford, UK: Oxford University Press.

Laible, D., & Song, J. (2006). Affect and discourse in mother–child co-constructions: Constructing emotional and relational understanding. *Merrill-Palmer Quarterly, 52*, 44–69.

Landry, S. H., Smith, K. E., & Swank, P. R. (2006). Responsive parenting: Establishing early foundations for social communication and independent problem-solving skills. *Developmental Psychology, 42*, 627–642.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.

Lay, K., Waters, E., & Park, K. A. (1989). Maternal responsiveness and child compliance: The role of mood as a mediator. *Child Development, 60*, 1405–1411.

Leerkes, E. M., Blankson, N., & O'Brian, M. (2009). Differential effects of maternal sensitivity to infant distress and non-distress on social-emotional functioning. *Child Development, 80*, 762–775.

Lepper, M. (1983). Social control processes, attributions of motivation, and the internalization of social values. In E. T. Higgins, D. N. Ruble, & W. W. Hartup (Eds.), *Social cognition and social development: A sociocultural perspective* (pp. 294–330). New York: Cambridge University Press.

Leve, L. D., & Fagot, B. I. (1997). Prediction of positive peer relations from observed parent–child interactions. *Social Development, 6*, 254–269.

Londerville, S., & Main, M. (1981). Security of attachment, compliance, and maternal training methods in the second year of life. *Developmental Psychology, 17*, 289–299.

Maccoby, E. E. (1999). The uniqueness of the parent–child relationship. In W. A. Collins & B. Laursen (Eds.), *Min-*

nesota symposia on child psychology: Vol. 30. Relationships as developmental contexts (pp. 157–175). Mahwah, NJ: Erlbaum.

Maccoby, E. E. (2007). Historical overview of socialization research and theory. In J. E. Grusec & P. D. Hastings (Eds.), *Handbook of socialization* (pp. 13–41). New York: Guilford.

Maccoby, E. E., & Martin, J. A. (1983). Socialization in the context of the family: Parent–child interaction. In E. M. Hetherington (Ed.), *Handbook of child psychology* (Vol. 4, 4th ed., pp. 1–102). New York: Wiley.

Matas, L., Arend, R., & Sroufe, L. A. (1978). Continuity of adaptation in the second year: The relationship between quality of attachment and later competence. *Child Development*, *49*, 547–556.

Mattanah, J. F., Pratt, M. W., Cowan, P. A., & Cowan, C. P. (2005). Authoritative parenting, parental scaffolding of long-division mathematics, and children's academic competence in fourth grade. *Applied Developmental Psychology*, *26*, 85–106.

McElwain, N. L., & Booth-LaForce, C. (2006). Maternal sensitivity to infant distress and nondistress as predictors of infant–mother attachment security. *Journal of Family Psychology*, *20*, 247–255.

McNamara, P., & Humphry, R. (2007). Now this is what you do: Developing structured routines. *OTJR: Occupation, Participation and Health*, *27*, 88s–89s.

Miller, P. J., & Goodnow, J. J. (1995). Cultural practices: Toward an integration of culture and development. *New Directions for Child and Adolescent Development*, *67*, 5–16.

Mischel, W., & Shoda, Y. (1995). A cognitive-affective-system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychology Review*, *102*, 246–268.

Mulvaney, M. K., McCartney, K., Bub, K. L., & Marshall, N. L. (2006). Determinants of dyadic scaffolding and cognitive outcomes in first graders. *Parenting: Science and Practice*, *6*, 297–320.

Neizel, C., & Stright, A. D. (2003). Mothers' scaffolding of children's problem solving: Establishing a foundation of academic self-regulatory competence. *Journal of Family Psychology*, *17*, 147–159.

Odden, H., & Rochat, P. (2005). Observational learning and enculturation. *Educational and Child Development*, *21*, 39–50.

Parpal, M., & Maccoby, E. E. (1985). Maternal responsiveness and subsequent child compliance. *Child Development*, *56*, 1326–1334.

Patterson, G. R. (1982). *Coercive family process*. Eugene, OR: Castalia Press.

Patterson, G. R., & Fisher, P. A. (2002). Recent developments in our understanding of parenting, bidirectional effects, causal models, and the search for parsimony. In M. H. Bornstein (Ed.), *Handbook of parenting* (Vol. 3, pp. 59–88). Mahwah, NJ: Erlbaum.

Pettit, G. S., Bates, J. E., & Dodge, K. A. (1997). Supportive parenting, ecological context, and children's adjustment: A seven-year longitudinal study. *Child Development*, *68*, 908–923.

Puntambekar, S., & Hübscher, R. (2005). Tools for scaffolding students in a complex learning environment: What have we gained and what have we missed? *Educational Psychologist*, *40*, 1–12.

Roberts, W., & Strayer, J. (1987). Parents' responses to the emotional distress of their children: Relations with children's competence. *Developmental Psychology*, *23*, 415–422.

Robinson, C. C., Mandleco, B., Olsen, S. F., & Hart, C. H. (1995). Authoritative, authoritarian, and permissive parenting practices: Development of a new measure. *Psychological Reports*, *77*, 819–830.

Rodriguez, M. L., Avduk, O., Aber, J. L., Mischel, W., Sethi, A., & Shoda, Y. (2005). A contextual approach to the development of self-regulatory competencies: The role of maternal unresponsivity and toddlers' negative affect in stressful situations. *Social Development*, *14*, 136–157.

Rogoff, B., Moore, L., Najafi, B., Dexter, A., Correa-Chavez, M., & Solis, J. (2007). Children's development of cultural repertoires through participation in everyday routines and practices. In J. E. Grusec & P. D. Hastings (Eds.), *Handbook of socialization* (pp. 490–515). New York: Guilford.

Rohner, R. P. (2005). Parental Acceptance–Rejection Questionnaire (PARQ): Test manual. In R. P. Rohner & A Khaleque. (Eds.), *Handbook for the study of parental acceptance and rejection* (pp. 43–106). Storrs, CT: Rohner Research.

Schwartz, S. H. (1994). Beyond individualism/collectivism: New cultural dimensions of values. In U. Kim, H. C. Triandis, C. Kagitcibasi, S.-C. Choi, & G. Yoon (Eds.), *Individualism and collectivism: Theory, method, and applications* (pp. 85–119). Thousand Oaks, CA: Sage.

Sears, R. R. (1957). Identification as a form of behavioral development. In D. B. Harris (Ed.), *The concept of development* (pp. 149–161). Minneapolis: University of Minnesota Press.

Sears, R. R., Maccoby, E. E., & Levin, H. (1957). *Patterns of child rearing*. Evanston, IL: Row Peterson.

Shields, A., & Cicchetti, D. (1997). Emotion regulation among school-age children: The development and validation of a new criterion Q-Sort Scale. *Developmental Psychology*, *33*, 906–916.

Shipman, K. L., & Zeman, J. (2001). Socialization of children's emotion regulation in mother–child dyads: A developmental psychopathology perspective. *Development and Psychopathology*, *13*, 317–336.

Smetana, J. G. (1997). Parenting and the development of social knowledge reconceptualized: A social domain analysis. In J. E. Grusec & L. Kuczynski (Eds.), *Parenting and children's internalization of values* (pp. 162–192). New York: Wiley.

Spagnola, M., & Fiese, B. H. (2007). Family routines and rituals: A context for development in the lives of young children. *Infants and Young Children*, *20*, 284–299.

Sperber, D., & Hirschfeld, L. A. (2004). The cognitive foundations of cultural stability and diversity. *Trends in Cognitive Sciences, 8*, 40–46.

Stayton, D., Hogan, R., & Ainsworth, M. D. S. (1973). Infant obedience and maternal behavior: The origins of socialization reconsidered. *Child Development, 42*, 1057–1070.

Thompson, R. A. (1994). Emotion regulation: A theme in search of a definition. In N. A. Fox (Ed.), The development of emotion regulation: Biological and behavioral considerations. *Monographs of the Society for Research in Child Development, 59*, 25–52.

Thompson, R., Laible, D., & Ontai, L. (2003). Early understandings of emotion, morality, and self: Developing a working model. In R. V. Kail (Ed.), *Advances in child development and behavior* (Vol. 31, pp. 137–171). San Diego, CA: Academic Press.

Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge, UK: Cambridge University Press.

Turiel, E. (1998). The development of morality. In W. Damon (Series Ed.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (5th ed., pp. 863–932). New York: Wiley.

Turner, V. D., & Berkowitz, M. W. (2005). Scaffolding morality: Positioning a socio-cultural construct. *New Ideas in Psychology, 23*, 174–184.

Vandell, D. L., & Wilson, K. S. (1987). Infants' interactions with mother, sibling, and peer: Contrasts and relations between interaction systems. *Child Development, 58*, 176–186.

Van der Mark, I. L., Bakersmans-Kranenburg, M. J., & van IJzendoorn, M. H. (2002). The role of parenting, attachment, and temperamental fearfulness in the prediction of compliance in toddler girls. *British Journal of Developmental Psychology, 20*, 361–378.

van IJzendoorn, M. H., & Hubbard, F. O. A. (2000). Are infant crying and maternal responsiveness during the first year related to infant–mother attachment at 15 months? *Attachment and Human Development, 2*, 371–391.

Verschueren, K., Dossche, D., Marcoen, A., Mahieu, S., & Bakermans-Kranenburg, M. (2006). Attachment representations and discipline in mothers of young school children: An observation study. *Social Development, 15*, 659–675.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Weir, K., & Duveen, G. (1981). Further development and validation of the Prosocial Behavior Questionnaire for use by teachers. *Journal of Child Psychology and Psychiatry, 22*, 357–374.

Wolin, S. J., & Bennett, L. A. (1984). Family rituals. *Family Processes, 23*, 401–420.

Wood, D. J., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*, 89–100.

Child Development, May/June 2010, Volume 81, Number 3, Pages 710–714

Commentary and Challenges to Grusec and Davidov's Domain-Specific Approach

Judy Dunn
King's College London

J. E. Grusec and M. Davidov's article (this issue) about domains of parenting and their links with different aspects of childhood outcome raises both interesting questions and challenges. Four of these concerns are discussed in relation to early childhood. First is the issue of bidirectionality. Recent studies highlight the contribution of individual differences in children, and of dyadic measures that include both child and parent, to the quality of parent–child relationships. Second, the specification of domains can be problematic, given the evidence of overlap between domains, and of combinations of different domains in real-life interactions. Third, developmental issues need to be addressed, as do cultural and cohort influences. Finally, children need to be considered as family members who monitor interactions between other family members, learn from and influence these others. The challenge to their account posed by models of family environments that include effects not shared by siblings merit attention.

The call for us to clarify the links between different aspects of parents' interactions with their children and those children's outcome is a welcome one. To differentiate between the various features of the relationship between child and parent clearly deserves attention. All too often the parent–child relationship is characterized in broad brush strokes, to the neglect of many aspects of what can clearly be a complex relationship. And the developmental issue of how key features of parent–child relationships in early childhood may change as children grow up is also still not well understood. Grusec and Davidoff (this issue) set out clearly specific hypotheses about the domains of parenting that they view as key and then examine which aspects of parenting are related to different aspects of child outcome. They have done us a real service in setting out their hypotheses and in summarizing what evidence there is for their key domains of socialization and the hypothesized links with outcome. In this commentary, the discussion is organized around four general topics in early childhood: first, the bidirectionality of influences between child and parent; second, the specification of domains of socialization; third, developmental issues raised by their model; and fourth, the significance of children as family members.

Bidirectionality

First, a general point concerning bidirectional effects. Until recently most research into socialization and parenting has focused on what parents do with or say to their children, and their feelings about them, rather than on the dyadic *relationship* between parent and child (for a notable exception, see Laible & Thompson, 2007), or the significance of bidirectional patterns of influence within this relationship. The classical scheme outlined by Baumrind (1971), based on the two dimensions of parental acceptance versus rejection and controlling versus permissiveness, is an example that has had huge influence and has led to an accumulation of evidence linking parenting as measured in a unidirectional way to differences in children's outcome. But differences in parents' behavior with their children may well be a response to differences in the *children's* behavior, both concurrently and in the history of their interactions over the longer term. Hinde pointed out decades ago that no measure of a parent's behavior can be considered to be independent of the history of previous interaction with the child, just as no aspect of the child's behavior toward the parent can be assumed to be independent of the history of their interaction (see, e.g., Hinde, 1979, 1987). He has argued moreover that we should take account of the contribution of

---

Correspondence concerning this article should be addressed to Judy Dunn, Social, Genetic and Developmental Psychiatry Centrem, Institute of Psychiatry, King's College London, London, UK SE5 8AF. Electronic mail may be sent to judy.dunn@kcl.ac.uk.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8103-0003

individual differences in parent–child relationships (both parent-to-child and child-to-parent interactions) to the cultural level of analysis.

We have to take seriously the contribution of individual differences in children to the relationship between child and parent. This was initially brought to my attention in a longitudinal study in which we conducted extensive home observations of babies from birth over the first 5 years of life (Dunn, 1977), and one of the relational processes we studied was the responses of mothers to their children when they cried—a classic aspect of maternal sensitivity (Ainsworth, Blehar, Waters, & Wall, 1978). However, it was clear that to treat this response to crying as a measure of sensitivity, we had to also take account of how much the particular baby had been crying and how irritable or difficult to soothe he or she was. We found that if the baby was one who relatively rarely cried, *and even so the mother was slow to respond to the crying*, this measure of maternal response was indeed an aspect of maternal insensitivity that related to other behavior that could be characterized as insensitive. However, if the baby was one who cried a great deal in the early weeks, and was difficult to soothe, a mother who over time became slow to respond to crying was not in other aspects of interaction insensitive (Dunn & Richards, 1977).

In the last 5 years there have been a series of investigations of bidirectional associations of parent–child relationships, brought together, for instance, in 2008 in a special issue of the *Journal of Abnormal Child Psychology*. The studies differ in the samples studied and in the breadth of measures; however, the commentary in the special issue highlights that "the studies cohere in providing clear evidence of bidirectionality: Child and youth adjustment difficulties contribute to declines in parent–child relationship quality over time, just as negative parenting (and parental depression) contributes to increases in child and youth adjustment problems over time" (Pettitt and Arsiwalla, 2008, p. 715).

Grusec and Davidoff acknowledge the "important and active contribution" children make to their own socialization; a feature of their model is the emphasis on the child's interpretation of parental practices and the relevance of the parental practices to the current domain in which the child is acting. However, the emphasis of their model remains chiefly on parental impact. In the domain of control, for example, they comment on the parent's arsenal of resources and the imbalance of power between child and parent. But there is little discussion of *dyadic* measures of the relationship, such as the meshing and involvement of the two partners, their mutual responsiveness or warmth, or the connectedness of their communication. In contrast, Laible and Thompson (2007) discuss how a broad quality of the parent–child relationship such as warmth "enhances or diminishes young children's receptivity to the socialization initiatives of others and in this manner also affects the responsiveness of others to the child" (p. 183). Laible and Thompson usefully distinguish between relational processes that are under parental control, those that are under children's control, and those that are under dyadic control. In the last category, for example, the conversations between child and parent play a key role in early social and moral understanding and in the development of conscience (Thompson, 2006). The development of moral sensitivity is one area among many where focusing on children's contribution to family conversations provides a perspective on the growth of understanding that differs from that of the Grusec and Davidoff model and could usefully be incorporated in the model.

This development of moral sensitivity is also a clear area for distinguishing cultural differences in socialization. For instance, a comparison of preschool children growing up in Pennsylvania and those in Cambridge, England (Dunn & Brown, 1991) highlighted a number of differences in explicit and implicit messages about the social world that were evident in mother–child conversations in control situations in the two cultural groups. Notions of harm, responsibility, culpability, principles of positive justice, and excuses of intent were all in place with these preschoolers, and this understanding was central to the children's reasoning in the power politics of the family. The children were already participants in conversations about the social rules and roles of their world, and this was evident not only in their disputes but also in their jokes, stories, and pretend play (Dunn, 1988). The two cultural groups differed both in the topics of the prescriptions and in their form. In Pennsylvania, the rights of the individual were frequently stressed; in Cambridge, the emphasis was on harm to others, politeness, and appropriate behavior. In the Pennsylvanian families, prohibitory messages were addressed to a child in terms of his or her personal actions and were less likely than those in Cambridge to include an evaluative message. Mothers made reference to *individual responsibility* such as "You put your toys away—you got them out!" "You get off the table!" In contrast, attempts to control the English children were less often focused on the individual child and often made references to general principles, as in the comment made by a mother to a 3-year-old: "Tables are not for walking!"

The issue of dyadic measures of child and parent, and their role in the development of social understanding, has been persuasively argued for by Tomasello and his colleagues in the context of research on the significance of joint engagement and attention (Behne et al., 2008; Tomasello, Carpenter, Call, Behne, & Moll, 2005). They show that joint engagement is "a powerful way for very young infants to develop an understanding of other people, in particular to come to know what others have and have not experienced" (Behne et al., 2008). It is not immediately clear how this joint engagement would be assigned to one of the five domains of socialization specified by Grusec and Davidoff—the second issue we consider.

Specification of Domains of Socialization

The notion of domains has been particularly developed in relation to cognitive development: Information from different content areas is suggested to be processed by specific modules, each of which is responsible for the acquisition and regulation of different features of knowledge (Keil & Wilson, 1999). Grusec and Davidoff apply this approach to socialization, with the proposal that each domain of the parent–child relationship is key for a particular aspect of socialization.

The question is: How many domains should be specified? Grusec and Davidoff's typology is fivefold: The domains are protection, reciprocity, control, guided learning, and group participation. In each of these domains, it is argued, different parental behavior is required for "effective socialization." Each domain is presented as prototypic. Yet as the authors note, in real-life interactions there are often combinations or overlaps among the domains. Both the advantages of such a typology (such as clarity and the opportunity to test hypotheses for evidence) and its disadvantages (how to deal with the evidence of overlap between domains in real-life interactions, for instance, and how to relate the account to other theoretical specifications of domains) are evident. Different theorists focus on different domains, and it is not simple to test the empirical basis for these different domains.

The specification of domains leads to some claims that appear inconsistent. In the protection/security domain for instance, Grusec and Davidoff view aspects of positive parenting such as reciprocity, appropriate discipline, and teaching/guided learning as *not* contributing to children's feelings of security. In the reciprocity domain, they argue that responsiveness to distress "should not facilitate reciprocated cooperation." In the control domain there are also similar arguments put forward by Grusec and Davidoff about what parental behavior is *not* effective in controlling children—and again these include some inconsistencies that may surprise the general reader. Responsiveness to distress, reciprocity, or facilitation of group participation are not viewed as processes that contribute to control. One wonders why distracting the child by play or conversation, or through humor should be excluded as a control strategy.

Developmental Issues

It would be useful to have data that could clarify how the prominence of different domains and how relational processes *within* domains may change as children grow up. In the 1980s, in the longitudinal study in Cambridge referred to above the children were growing up in families that differed widely in social class, neighborhood, and parental education. We were particularly interested in early individual differences in child–parent interaction, and during the extensive observations made over the babies' first 14 months, individual differences in measures of the mothers' responsiveness to their babies' distress and crying were remarkably stable. This maternal responsiveness to crying in the 1st year was not related to the mothers' educational level or social class. However during the 2nd year, as the children began to understand and use language, a different pattern of individual differences in maternal sensitivity emerged. The stable pattern of responsiveness to distress and crying disappeared, and a new ranking of individual differences in responsiveness to the babies' attempts to communicate emerged, which was clearly related to the mothers' educational level.

Developmental changes were also evident in gender differences and social class differences in parent–child and child–sibling interactions as the children reached middle childhood. These were particularly apparent among the working-class boys, who spent more time out of the family home and on the streets. They were already active members of their peer groups, well before the boys in middle-class families, who spent more time with their parents. Similar social class differences were found at the time in a study of 700 children by Newson and Newson (1976). It is quite likely that such gender and social class differences in peer group associations are less evident today than they were in the 1970s, a note of caution that reminds us that cultural differences in social relationships change over time. For a theoretical model of parenting that has relevance today, the Grusec and Davidoff approach needs to address both the issue of developmental change and the issue of cultural influence, including possible changes in such cultural influences over the decades.

Children as Family Members

The fourth issue raised by Grusec and Davidoff's article concerns children as family members and the complexity of influences on their socialization. From the end of the 1st year, there is evidence for influences on children that involve people other than the mother, father, or chief caretaker. For instance, individual differences in children's early understanding of others' feelings, thoughts, beliefs, intentions—their theory of mind—(noted by Grusec and Davidoff to be key in socialization) are related not only to their participation in conversations about these issues with family members but also to what is happening between their parents and siblings. Both experimental studies and naturalistic observational research have shown the powerful effect on children of witnessing angry interactions between others (Cummings, Zahn-Waxler, & Radke-Yarrow, 1987). Within the family, children are vigilant monitors of the interactions between their parents and siblings. In a study involving naturalistic home observations in families in Cambridge, the children's behavior as 18-month-olds and as 24-month-olds when disputes took place between their mother and older sibling was monitored (Dunn, 1988). These disputes were of much interest to the second-born children. They rarely ignored them but responded to 72% of the disputes at 18 months and to 78% at 24 months. The ways in which they responded varied. In some incidents they simply watched, soberly; in others they imitated the older sibling's prohibited action and laughed. Sometimes they attempted to support one or other of the antagonists. Our analyses showed that the emotion expressed by the disputants was important but the topic on which the disputes were focused was also key (Dunn, 1988). The mastery of cultural matters that takes place during the preschool years evidently reflects the interest of the child as a family member, who is motivated to understand the social ways and prescriptions of her world in part because of the pressure of her relationships within that world.

In families in which there were relatively frequent disputes about control issues between parent and older sibling, the younger siblings' understanding of mind was especially evident over the following months (Dunn, Brown, Slomkowski, Tesla, & Youngblade, 1991). The analysis of which factors were linked to these individual differences in mind reading highlighted the significance of three quite different sets of social processes. One was the extent of mother–child discourse about why people behave the way they do, their feelings and mental states, in which children were involved. The second was the extent of disputes between mother and older sibling that the second-born child had witnessed. And the third was the extent to which the second-born child had been involved in pretend play with the firstborn. There is evidence that over the 3rd and 4th years, cooperative pretend play with an older sibling, and especially with a friend, provides contexts that are especially rich in conversations that involve discussion of other minds, and of why people behave the way they do. Shared pretense with another child, it has been argued by Tomasello and colleagues, is one of the earliest areas in which children enter into "collective intentionality" (Behne et al., 2008). They have also shown, with experimental studies, that explicit discourse about pretending is crucial in the development of reflective understanding of pretense (Tomasello, 1999). Including family members, such as siblings, in analyses of domain approaches for socialization would be one way to construct a more interactive model of parenting. For example, how do siblings respond to parental strategies aimed at another member of the family (such as a sibling) and in what way does the parental strategy get modified when directed to a second-born child, who has witnessed the parent's interactions with the firstborn? Including these analyses would provide a nuanced and full picture of socialization from a domain approach.

In summary, we should be grateful to Grusec and Davidoff for framing their ideas on socialization in systematic fashion. They set out helpfully both the differences and the connections between their model and that of Turiel's (1983) social domain theory. Their account is full of interest and makes many novel points, and it is provocative in stimulating others to reflect on how far and in what ways their own ideas mesh with the model outlined, and how far they differ. The challenges now are many: for instance, to investigate how far the study of other cultures supports their views, and how appropriate their approach is for the new models of family environments such as those stimulated by awareness of how siblings within the same family differ (Plomin, Asbury, & Dunn, 2001).

References

Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. (1978). *Patterns of attachment*. Hillsdale, NJ: Erlbaum.

Baumrind, D. (1971). Current patterns of parental authority. *Developmental Psychology, 4*, 1–103.

Behne, T., Carpenter, M., Grafenhain, M., Liebal, K., Liskowski, U., Moll, H., et al. (2008). Cultural learning and cultural creation. In U. Mueller, J. I. M. Carpendale, N. Budwig, & B. Sokol (Eds.), *Social life and social knowledge* (pp. 65–101). Mahwah, NJ: Erlbaum.

Cummings, E. M., Zahn-Waxler, C., & Radke-Yarrow, M. (1987). Coping with background anger. *Child Development, 85*, 976–984.

Dunn, J. (1977). Patterns of early interaction: Continuities and consequences. In H. R. Schaffer (Ed.), *Studies in parent-infant interaction* (pp. 457–474). London: Academic Press.

Dunn, J. (1988). *The beginnings of social understanding*. London: Blackwell.

Dunn, J., & Brown, J. (1991). Becoming American or English? Talking about the social world in England and the United States. In M. H. Bornstein (Ed.), *Cultural approaches to parenting* (pp. 155–172). Hillsdale, NJ: Erlbaum.

Dunn, J., Brown, J., Slomkowski, C., Tesla, C., & Youngblade, L. (1991). Young children's understanding of other people's feelings and beliefs: Individual differences and their antecedents. *Child Development, 62*, 1352–1366.

Dunn, J., & Richards, M. P. M. (1977). Observations on the developing relationship between mother and baby in the newborn period. In H. R. Shaffer (Ed.), *Studies in parent-infant interaction* (pp. 427–455). London: Academic Press.

Grusec, J. E., & Davidov, M. (2010). Integrating different perspectives on socialization theory and research: A domain-specific approach. *Child Development, 81*, 687–709.

Hinde, R. A. (1979) *Towards understanding relationships*. London: Academic Press.

Hinde, R. A. (1987). *Individuals, relationships, and culture*. Cambridge, UK: Cambridge University Press.

Keil, F. C., & Wilson, R. A. (Eds.). (1999). *Explanation and cognition*. Cambridge, MA: MIT Press.

Laible, D., & Thompson, R. A. (2007). Early socialization: A relationship perspective. In J. E. Grusec & P. D. Hastings (Eds.), *Handbook of socialization: Theory and research* (pp. 181–207). New York: Guilford.

Newson, J., & Newson, E. (1976). *Seven years old in the home environment*. London: Allen & Unwin.

Pettitt, G. S., & Arsiwalla, D. D. (2008). Commentary on special section on bidirectional parent-child relationships: The continuing evolution of dynamic, transactional models of parenting and youth behaviour problems. *Journal of Abnormal Child Psychology, 36*, 711–718.

Plomin, R., Asbury, K., & Dunn, J. (2001). Why are children in the same family so different? Nonshared environment a decade later. *Canadian Journal of Psychiatry, 46*, 225–233.

Thompson, R. A. (2006). The development of the person: Social understanding, relationships, self, conscience. In W. Damon & R. Lerner (Series Eds.), & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (6th ed., pp. 24–98). Hoboken, NJ: Wiley.

Tomasello, M. (1999). *The cultural origins of cognition*. Cambridge, MA: Harvard University Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences, 28*, 675–735.

Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge, UK: Cambridge University Press.

Child Development, May/June 2010, Volume 81, Number 3, Pages 715–719

Modules, Theories, or Islands of Expertise? Domain Specificity in Socialization

Susan A. Gelman
University of Michigan

The domain-specific approach to socialization processes presented by J. E. Grusec and M. Davidov (this issue) provides a compelling framework for integrating and interpreting a large and disparate body of research findings, and it generates a wealth of testable new hypotheses. At the same time, it introduces core theoretical questions regarding the nature of social interactions, from the perspective of both children and their caregivers. This commentary draws on the literature regarding domain specificity in cognitive development, applauds what is innovative and exciting about applying a domain-specific approach to socialization processes, and points to questions for future research. Foremost among these is what is meant by ''domain specificity.''

Grusec and Davidov (this volume) have provided us with a valuable and thought-provoking framework for thinking about socialization, arguing in compelling detail for the importance of domain-specific processes. They lucidly argue that parent–child relationships and mechanisms have been obscured by treating socialization processes as undifferentiated, and provide numerous intriguing examples of distinctions among their hypothesized five domains: protection, reciprocity, control, guided learning, and group participation. It is misleading, for example, to talk about parental ''sensitivity'' to the child, as the nature and focus of that sensitivity differs, depending on whether one is focused on (say) guided learning or reciprocity. When one talks about sensitivity writ large, the evidence is contradictory, the theories are competing, and the correlations are low. Yet when one considers sensitivity from the perspective of distinct domains, seeming inconsistencies are only apparent, and structure emerges at the domain level. Somewhat paradoxically, Grusec and Davidov suggest that the diverse approaches to socialization discussed in the literature can be unified *only* when one accepts that socialization is not a unitary phenomenon.

It is particularly novel and exciting to see the concept of domain specificity extended from the realm of cognition to that of socialization. We have known for decades that cognitive processes and cognitive development are uneven across domains (Hirschfeld & Gelman, 1994). Children are much worse than adults at learning calculus but much better than adults at learning syntax. A child chess expert shows better recall of the placement of chess pieces on a chessboard than a chess-novice adult—while simultaneously performing worse if the recall task includes numbers rather than words. Although thoughtful domain-differentiated models of socialization have also been offered previously (see Bugental, 2000; Smetana, 2006; Turiel, 1983, 2006, for notable examples), Grusec and Davidov imply that domain-general models continue to be dominant. To the extent that domain-general models of socialization have persisted, whereas domain-specific models of cognition (including social cognition) have received more general acceptance, this difference may reveal something interesting about historical assumptions about the nature of socialization and cognitive development.

Advantages to a Domain-Specific Approach

There are obvious benefits of a successful domain-specific perspective. In addition to bringing unity

---

Preparation of this commentary was supported by NICHD Grant R01 HD36043. I thank Melanie Killen for her detailed and thoughtful comments on an earlier draft.

Correspondence concerning this article should be addressed to Susan A. Gelman, 530 Church St., Department of Psychology, University of Michigan, Ann Arbor, MI 48109-1043. Electronic mail may be sent to gelman@umich.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc. All rights reserved. 0009-3920/2010/8103-0004

to what otherwise seems like a dizzying array of approaches, the domain-specific framework possesses the hallmark of a good scientific theory: It is predictive. Many precise, testable claims emerge, as delineated in Grusec and Davidov. For example, children who are securely attached (due to parental sensitivity within the protection domain) are predicted to be more likely to comply in situations that involve their own safety (and thus fall within the protection domain). Furthermore, prosocial behaviors are predicted to vary by domain (e.g., responding to distress vs. friendliness and cooperation vs. inhibiting one's self-interest).

Grusec and Davidov's approach also has the virtue of generativity. Reading through the article, one cannot help but speculate about connections to other issues. Some future research questions might include: Are there stable individual differences with respect to which domain is dominant in a given individual? How might factors such as gender or temperament influence a parent's ease or skill in operating in any particular domain? Within a family system, do different individuals tend to meet different domain needs of the child—with mother, father, sibling, teacher each having a "key" domain? Is there cultural variation in terms of which domains are more consistently valued? How do family-based, childhood experiences within a domain influence later relationships? Which of the domains identified by Grusec and Davidov are found in nonhuman primates or other social organisms, and what does this reveal about human societies?

Another aspect of the approach that is very appealing is that the child's contribution to the social interaction is central. As stated by Grusec and Davidov, "Children are important and active contributors to their own socialization" (p. 692). They also note that the meaning of the behavior must be considered from the child's point of view and that the child's causal construal may be critical (e.g., the very same display of affection can convey protection or group participation). I wish to underscore this point, because it is one that is all too often missed. When focusing on parental contributions to development, a potential danger is that the child is viewed as a passive consumer of adult input—the environment does all the work, and the child is a mere recipient. Conversely, in the cognitive development literature, the emphasis has been on the "child as lone scientist," figuring out the world in the absence of others that support the child's learning, pass knowledge down generations, supply cultural tools, and so on. The domain approach recognizes—indeed, requires—that environmental inputs be matched to child capacities and expectations. As obvious as it may seem, there has been a recent turn to acknowledging the importance of the interaction of the child and others that support the child's learning and development (e.g., Csibra & Gergely, 2009; Gelman, 2009; Harris & Koenig, 2006). Innate skeletal frameworks go hand in hand with learning from experience (Gelman & Williams, 1998); parental input and child biases work together toward a common goal (Markman, 1992),

A theoretically rich proposal, such as that offered by Grusec and Davidov, invariably raises many questions. I will discuss two questions here: What is a domain? And, what is the relation among domains? These questions undoubtedly reflect my own background as a cognitive developmental psychologist with strong interests in the role of parent–child interactions in concept development (see, e.g., Gelman, 2009).

What Is a Domain?

In the field of cognitive development, there have been at least three distinct approaches to domain specificity, with distinctly different claims regarding developmental mechanisms, what is innately specified, how variable the outcome can be, and the role of input (see Hirschfeld & Gelman, 1994; Wellman & Gelman, 1998, for a review). Domain specificity can reflect *modularity*, in which domains have emerged through evolutionary pressures, have innate systems for taking in and processing specific sorts of information, and represent biologically constrained processes (examples include vision, natural language syntax, number, and face recognition). Modular domains are universal, neurologically localized and acquired in the presence of minimal cues; they unfold on a strict maturational timetable; and they experience sensitive periods (Cosmides & Tooby, 2001; Scholl & Leslie, 1999). Domain specificity can also reflect *expertise*, in which domain-specific abilities are highly variable across individuals and emerge through extensive experience (examples include expertise in chess or reading, which profoundly changes the processing capacities of the individuals who engage in these experiences—but only for the domain in question). Expertise-based domain specificity is not innately specified (Chi, 1978; Ericsson, 2005). Finally, domain specificity can reflect *causal theories*. The child starts with skeletal domain-specific principles and ontologies (examples include theory of mind or theory of physical objects), which are then informed by data and evidence so that the child constructs increasingly complex and detailed causal-explanatory understandings (Carey, 2009; Gopnik & Meltzoff, 1997).

Although Grusec and Davidov suggest that socialization domains are modular (''specialized modules, each of which has evolved to accomplish a recurring important task or function [e.g., face recognition, numerical thinking, grammar processing]''; p. 687), at times the socialization domains also appar to have features of theories or expertise, respectively. They are theory-like in incorporating causal explanatory reasoning. For example, in order to discern whether a prosocial behavior is mere politeness or true empathy, one must ask the child *why* he has behaved as he did—in other words, to assess the causal understanding. Similarly, Johnson, Dweck, and Chen (2007) present fascinating experimental evidence that infants with different attachment types possess different causal theories of social interaction, as assessed in their reasoning about third-party characters. Furthermore, the possibility of individual or cultural influences on the salience of different domains suggests that expertise may play an important role. Thus, an interesting empirical question is to examine in more detail some of the features of modular theories (biological foundation, universality, etc.) to see how well they fit the various domains of socialization.

Of course, it is even possible that protection, reciprocity, control, and so on are not, strictly speaking, domains at all but instead ''types''—of relationships, goals, functions, values, or attentional foci. A widely recognized concern in the cognitive literature is that the notion of ''domain'' can be multiplied to include nearly any perceived distinction. Keil (2000) notes: ''It seems more than any pattern of information in the world for which the body has a specialization. Surely innate allergies to poison ivy do not mean we have domain specific constraints on learning about poison ivy, even if those allergies result in all of us learning about poison ivy differently from daisies.'' Similarly, not all differences in social processes can or should rise to the level of ''domain.'' As an example, a review article by Medin, Lynch, and Solomon (2000) elegantly lays out the importance of considering distinctions among different kinds of concepts. They review structural differences, processing differences, and content-laden differences, and make a convincing case that we ignore such variation at our peril. Yet these different *types* of concepts are not different *domains* of concepts.

To be clear, posing the question of whether these domains are modular, or indeed whether they are properly considered domains, does not in any way diminish the value and importance of the typology laid forth in Grusec and Davidov. Rather, the question is to take seriously the notion of ''domain'' and to work through, in systematic fashion, what the implications are of such a view. The innovative approach of dividing up socialization behaviors into distinct types is an important start to what promises to be a rich program of research.

What Is the Relation Among Domains?

One recurring (and fascinating) theme in the article is that the fit between maternal goals and child goals—or lack of fit—has important consequences. Thus, a mother who construes a situation as falling within the control domain, and thus exerts discipline, will miss the mark if her child construes the situation as falling within the protection domain and seeks comfort and soothing. The proposal here seems to suggest that an interaction is largely centered within a particular domain (e.g., control vs. protection, as in the example above), both from the standpoint of the parental input and from the standpoint of the child uptake. This proposal, although plausible, is actually an open empirical question and depends on the relation of one domain to another.

I illustrate the problem with an analogy from considering cognitive domains, specifically, domain-specific causal theories. Consider the hypothetical case of a child who asks why she is not allowed to eat all her Halloween candy in one night. A parent might respond: ''Because I think too much candy will make you sick.'' The causal structure of this information can be firmly located within a domain, in this case biological, as it concerns a nonintentional, natural bodily process: illness. The parent instead could have explained the very same restriction by referring to an internal mental state such as desire (''Because you're going to want to eat more candy tomorrow''), which would place the behavior in a psychological domain. Or, the parent could have explained the restriction by referring to deontic rights or obligations, or appealing to ethical principles (''It's wrong to be greedy, and besides, you need to share your candy with your sister''), which would place the behavior in a moral domain. And so forth. Indeed, much of human activity can be understood within multiple domains, because humans themselves are simultaneously physical, psychological, biological, even spiritual beings. Yet in providing an explanation, the speaker makes a choice as to which causal framework is most relevant. In expressing an explanation, one is forced to choose a particular causal story, located within a particular domain. In this example, the parent selects among domains, representing a particular perspective, and one could well imagine (say) a parent–child mismatch in this regard (e.g., if the child is focused on a psychological rationale [''I want the candy''], whereas the parent focuses on the biological rationale; see Smetana, 1989, for mismatches of this sort in parent–adolescent conflicts).

However, in another sense, a single explanation can be rich with multiple signals that are available to the child, across domains, revealing not only a particular causal framework but also informational cues regarding syntax, speech perception, pragmatics, phonetics, conceptual content, and theory of mind. A single sentence (e.g., ''Because I think too much candy will make you sick'') informs the child about: the biological process of illness, psychological states (''I think. . .''), syntactic constructions in English (the embedded clause), generic nouns (''candy''), parental values, regional accent, and so on. Various kinds and levels of information are bundled into a single, multivalent signal. A child may be interested in talking about candy but is also exposed to and learning about many other things simultaneously. Spelke (2003) has argued that human language is in fact unique among communication systems in permitting the juxtaposition of thought from multiple domains. Whereas a rat can successfully represent both geometric (shape) and nongeometric (color) information about a large-scale space (e.g., a room), only a verbal human can combine the geometric and nongeometric information into a single representation. We have ''a uniquely human combinatorial capacity that serves to conjoin [core knowledge] representations to create new systems of knowledge'' (Spelke, 2003, p. 305).

The question then is, in the realm of socialization, what is the nature of the social signal to the child, in terms of what the parent implies and what the child takes in? Is it more like the first case, in which an interaction is typically located within a single domain, or is it more like the second case, in which multiple strands of signal are available simultaneously? For example, even if a parent intends to control the child and place constraints on the child's behavior, she may simultaneously be sending powerful signals regarding group participation via social-conventional reasoning (''this is the sort of behavior that we do, as members of group X''; see Killen, Sinno, & Margie, 2007), guided learning (teaching about kinds of things in the world, such as candy; see Csibra & Gergely, 2009, for the suggestion that this ''pedagogical'' stance may be uniquely human), even protection (''I don't want you to get sick'') and reciprocity (''I'll let you have one more piece of candy, though, since you asked so nicely''). Grusec and Davidov's framework certainly allows for multiple signals of this sort; in their article they explicitly acknowledge that different domains are highly interconnected. Moreover, in the realm of guided learning in particular, a great deal of implicit instruction takes place in everyday interactions that on the surface may not appear to be pedagogical—occasions in which a child is simply engaged in conversation, or observing another's actions (e.g., Gelman, Taylor, & Nguyen, 2004; Meltzoff, 2002; Tomasello, Carpenter, Call, Behne, & Moll, 2005). Thus, the question is not *whether* domains interact. Instead, the question may be more aptly framed as, which is ''standard operating procedure'': pursuit of a single domain within an interaction or the bundling of messages (both explicit and implicit) in a single interaction?

Conclusions

There is a rich history of domain-specific approaches to cognition, but these approaches are relatively less often applied to social interactions. The approach pioneered by Grusec and Davidov (building on related though distinct work by Bugental, 2000; Smetana, 2006; Turiel, 1983, 2006, among others) raises an important set of questions regarding the structure and process of socialization. Future work would benefit by considering: (a) whether these interaction types or goals are most accurately considered modules, theories, or islands of expertise, and (b) the extent to which the domains overlap or present simultaneously within a social interaction.

References

Bugental, D. B. (2000). Acquisition of the algorithms of social life: A domain-based approach. *Psychological Bulletin, 26*, 187–209.

Carey, S. (2009). *The origin of concepts*. New York: Oxford University Press.

Chi, M. T. H. (1978). Knowledge structures and memory development. In R. Siegler (Ed.), *Children's thinking: What develops?* (pp. 73–96). Hillsdale, NJ: Erlbaum.

Cosmides, L., & Tooby, J. (2001). Unraveling the enigma of human intelligence: Evolutionary psychology and the multimodular mind. In R. J. Sternberg & J. C. Kaufman (Eds.), *The evolution of intelligence* (pp. 145–198). Mahwah, NJ: Erlbaum.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*, 148–153.

Ericsson, K. A. (2005). Recent advances in expertise research: A commentary on the contributions to the special issue. *Applied Cognitive Psychology, 19*, 233–241.

Gelman, S. A. (2009). Learning from others: Children's construction of concepts. *Annual Review of Psychology, 60*, 115–140.

Gelman, S. A., Taylor, M. G., & Nguyen, S. (2004). Mother-child conversations about gender: Understanding the acquisition of essentialist beliefs. *Monographs of the Society for Research in Child Development, 69*(1, Serial No. 275).

Gelman, R., & Williams, E. (1998). Enabling constraints for cognitive development and learning: Domain specificity and epigenesis. In D. Kuhn & R. Siegler, (Eds.). *Cognition, perception and language: Vol. 2. Handbook of child psychology* (5th ed., pp. 575–630). New York: Wiley.

Gopnik, A., & Meltzoff, A. (1997). *Words, thoughts, and theories*. Cambridge, MA: MIT Press.

Grusec, J. E., & Davidov, M. (2010). Integrating different perspectives on socialization theory and research: A domain-specific approach. *Child Development, 81*, 687–709.

Harris, P. L., & Koenig, M. A. (2006). Trust in testimony: How children learn about science and religion. *Child Development, 77*, 505–524.

Hirschfeld, L. A., & Gelman, S. A. (1994). Toward a topography of mind: An introduction to domain specificity. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 3–35). New York: Cambridge University Press.

Johnson, S., Dweck, C. S., & Chen, F. (2007). Evidence for infants' internal working models of attachment. *Psychological Science, 18*, 501–502.

Keil, F. C. (2000). Nurturing nativism, review of Cowie's *What's within*. In *A field guide to the philosophy of mind*. Retrieved March 18, 2010, from http://host.uniroma3.it/progetti/kant/field/cowiesymp_keil.htm

Killen, M., Sinno, S., & Margie, N. G. (2007). Children's experiences and judgments about group exclusion and inclusion. In R. Kail (Ed.), *Advances in child psychology* (Vol. 35, pp. 173–218). New York: Elsevier.

Markman, E. M. (1992). Constraints on word learning: Speculations about their nature, origins, and domain specificity. In M. R. Gunnar & M. Maratsos (Eds.), *Modularity and constraints in language and cognition: Vol. 25. The Minnesota symposia on child psychology* (pp. 59–101). Hillsdale, NJ: Erlbaum.

Medin, D. L., Lynch, E. B., & Solomon, K. O. (2000). Are there kinds of concepts? *Annual Review of Psychology, 51*, 121–147.

Meltzoff, A. N. (2002). Imitation as a mechanism of social cognition: Origins of empathy, theory of mind, and the representation of action. In U. Goshwami (Ed.), *Handbook of childhood cognitive development* (pp. 6–25). Oxford, UK: Blackwell.

Scholl, B. J., & Leslie, A. M. (1999). Modularity, development, and "theory of mind." *Mind & Language, 14*, 131–153.

Smetana, J. G. (1989). Adolescents' and parents' reasoning about actual family conflict. *Child Development, 60*, 1052–1067.

Smetana, J. G. (2006). Social domain theory: Consistencies and variations in children's moral and social judgments. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 119–154). Mahwah, NJ: Erlbaum.

Spelke, E. S. (2003). What makes us smart? Core knowledge and natural language. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the investigation of language and thought* (pp. 277–311). Cambridge, MA: MIT Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences, 28*, 675–735.

Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge, UK: Cambridge University Press.

Turiel, E. (2006). The development of morality. In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (6th ed., pp. 789–857). New York: Wiley.

Wellman, H. M., & Gelman, S. A. (1998). Knowledge acquisition. In D. Kuhn & R. Siegler (Eds.), *Handbook of child psychology: Vol. 2: Cognitive development* (5th ed., pp. 523–573). New York: Wiley.

Child Development, May/June 2010, Volume 81, Number 3, Pages 720–726

Domain Specificity in Social Interactions, Social Thought, and Social Development

Elliot Turiel
University of California, Berkeley

J. E. Grusec and M. Davidov (this issue) have taken good steps in formulating a domain-specific view of parent–child interactions. This commentary supports the introduction of domain specificity to analyses of parenting. Their formulation is an advance over formulations that characterized parental practices globally. This commentary calls for inclusion of definitions of the classification system of domain-specific interactions and criteria for each domain. It is also maintained that Grusec and Davidov's domains of social interaction imply that processes of development are involved, along with socialization; that bidirectionality in parent–child relations needs to be extended to include mutual influences and the construction of domains of social thought; and that conflicts and opposition within families coexist with compliance and social harmony.

One of the mainstays in developmental psychology is the study of families and, especially, the influences of parents on children's development. The long-standing and enduring concern with families among developmental researchers hardly needs justification. Children spend a great deal of time in family interactions and most caregivers in most cultures take major responsibility for the care, protection, nurturance, and guidance of their children in many of life's realms. Because of these features—that children spend so much time, day in day out, with their parents and parental guidance encompasses so many realms—understanding family influences on children's development is a difficult and complex endeavor. It is a difficult and complex endeavor because so much goes on in family life over a period of many years during which children's thought and emotions are changing, their worlds outside the family are expanding, and their relationships with parents are undergoing transformations (Smetana & Turiel, 2003). And with time parents' ideas about children and parenting can change through their experiences as parents (Cowan, Powell, & Cowan, 1998; Sigel & McGillicuddy-DeLisi, 2002). Similarly, children's reactions to their parents can change with time as a consequence of their experiences with parents (Smetana & Turiel, 2003).

Adding even more to the difficulty of studying parenting and development is that the influences of many other activities on both parents and children need to be coordinated with direct family influences (Dunn, 1988). Parents engage in many activities in and out of the family that might bear on their parenting. These include parents' relationships with each other, their work and the financial needs of the family (Crouter & McHale, 2005); their efforts at achieving personal goals in life (and the conflicts that might be involved with goals for their children; Nucci, 2001); and their friendships and recreational activities—and, no doubt, much more. Children, too, engage in many activities in and out of the family that could have a bearing on how they are influenced by parents. These include relationships with siblings and other significant family members (Dunn, 1988), with friends and acquaintances (Rubin, Bukowski, & Parker, 2006), and with teachers, as well as their schooling, recreational activities, and concerns with social networks—and, no doubt, much more.

Global or Domain Specific?

All of this is to say that family life is complicated. Given all these complexities, it is both paradoxical and unsurprising that research into parental influences on children's development has a history

---

Correspondence concerning this article should be addressed to Elliot Turiel, Graduate School of Education, University of California, Berkeley, CA 94720. Electronic mail may be sent to turiel@berkeley.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8103-0005

of framing parental practices in general or global ways (Bornstein, 2006; Turiel, 2005). Simplifying the complex is one of the means often used to make matters researchable and comprehensible (especially in a time when parsimony was seen as necessary in scientific explanation). As discussed in the historical overview provided by Grusec and Davidov (this issue), in early formulations general mechanisms of learning such as reinforcement, secondary drives, and modeling were proposed as a means of explaining how parental actions accounted for children's social acquisitions. These learning mechanisms were sometimes seen as ways of simplifying the theoretical formulation put forth by Freud (1923/1960), with its dynamic model of unconscious processes and maturational stages (Bandura & Walters, 1963). This was followed by propositions regarding general or global styles of parenting. These were meant to provide a little more specificity than general mechanisms of learning through categories meant to capture how parents relate to children. Although distinctions between parents were drawn by virtue of the different categories, a particular parent's style was characterized in global terms and as a general type (e.g., primarily involving power assertion *or* love withdrawal *or* induction; alternatively, involving styles that are authoritarian *or* authoritative *or* permissive).

In their bold and forward-looking proposals for a domain-specific approach to socialization, Grusec and Davidov provide needed correctives for some problems, which I list below, in the global accounts of parenting. In my view, the extant global characterizations of parenting and the associated formulations of children's development are limited because of the following shortcomings:

1. That parents are portrayed as usually employing one general orientation in raising and disciplining their children.
2. That the process of parent–child interactions is seen as one-directional, rather than bidirectional and reciprocal.
3. That parents are seen to have the major, overwhelming causal influence on children's development.
4. That it is mainly through parental practices of childrearing that children more successfully or less successfully become competent members of the community (society, culture) through compliance, conformity, accommodation, and adaptation to the group, as well as internalization of rules, rituals, and practices, and deference to authority.

The major corrective that Grusec and Davidov address is to the first problem I listed. They also address issues pertaining to one-directional or bidirectional processes, touch on the question of the major influences of parents, and largely do not address the question of socialization as compliance and accommodation.

Five Domains of Social Interactions: A Good Plan

Grusec and Davidov's domain-specific proposal is a good means for moving away from broad characterizations of parent–child interactions because it is meant to provide an integration of approaches heretofore fragmented, is bolstered by some findings from studies that were not designed to examine domain specificity, and is aimed at providing a framework for research that would be designed to address domain distinctions and verify their validity, and provides a "first step . . . before considering more complex combinations" (p. 691).

As Grusec and Davidov state (p. 701), "There is considerable initial evidence to support the contention that socialization processes differ in these domains, although additional research is obviously greatly needed." The categories they propose for different interactions between parents and children constitute an important advance over the general, broad, and unitary types of parenting previously emphasized in socialization research. In my view, the broad categories fail to capture parents' likely use of different practices for different situations and contexts and the ways they are likely to combine different types of practices.

More generally, it seems to me unlikely that parents would relate in one general fashion to their children regardless of the types of actions or desires on the part of their children in any given situation. There is a good deal of evidence in the psychological literature (developmental or otherwise) that people do not respond in the same ways across situations or contexts (Ross & Nisbett, 1991). There are debates about how to explain contextual variations in thought or action, but a good deal of evidence exists showing that people do not approach the world in a general way or through one type of orientation (Turiel, 1998). People draw priorities and make distinctions in choices and goals—distinctions that likely apply to parenting practices. As I have previously stated in discussing domains of social thought (Turiel, 2005, p. 80):

> Most parents are likely to respond much more strongly, assertively, and with greater efforts at control if a child engages in actions that pose a danger to others or to the self or that are illegal (for example, theft) than if the actions have to do with manners, how to dress, or which play and leisure activities to pursue. In addition, actions that have to do with harm or illegalities are likely to be treated differently from actions that have to do with children's educational pursuits and goals, even when parents regard education as very important and it is taken seriously. . . . Parents who think that a friend may lead their child in undesirable directions will react differently from the choice of a friend they simply do not like.

Grusec and Davidov provide a systematic analysis of the types of differences that they glean from existing research. One of the examples they provide nicely illustrates how global parenting variables are not always predictive of positive outcomes in children and that parents do not generally relate to children in one way. I am referring to the finding (p. 700) that parent scaffolding, but not warmth and control, predicted successful cognitive functioning. However, this example could be taken to mean that the categories are artificial or too restrictive because it might be that a combination of parental practices would best predict outcomes. In this example, it might be that a combination of scaffolding and warmth (and perhaps others) would be more effective in children's learning than one type. This is a question that can be put to empirical test. I believe, however, that Grusec and Davidov are correct in arguing that combinations and overlaps in domains do not invalidate the distinctiveness of the domain categories. If the domains identified are valid, it is necessary to identify and define them in order to understand interactions that involve combinations and overlap. Research on domains of moral, conventional, and personal understandings (referred to by Grusec and Davidov as social knowledge domains) has shown that many social decisions are best explained by assessing the distinct domains and how those are coordinated in situations that involve domain combinations (Killen, 1989; Turiel, 2008).

In discussing research into complex combinations, Grusec and Davidov (p. 691) state, "Outlining the unique and distinctive features of each domain seems an important first step." In my view, to conduct research to ascertain the validity of the domains and to use them for further research on socialization and development, it is necessary to go beyond outlining distinctive features of the domains by formulating systematic definitions of each domain and criteria for their distinctiveness (Chomsky, 1979; Piaget, 1970; Turiel, 1983, 1998). This is to say that researchers need to know a domain when they see one. The best way to know when you see it is to have a classification system that includes:

1. An overriding definition of parent–child interactions that encompasses, at least provisionally, the five domains, that is, the common features of the five domains that render them part of the same classification system. Such features should allow for classification of certain types of actions as domains within the system and other types that do not constitute domains. For instance, Grusec and Davidov do not classify responsiveness, warmth, and harshness as domains, in themselves, but as related to the five domains. An overarching definition of the system of classification would provide a basis for including, say, protection and control as domains and not responsiveness, warmth, and harshness, and
2. Definitions of each of the domains including criteria for what constitutes each domain and distinguishes one from another. Ideally, this would take the form of a coding system to be applied reliably in analyses of parent–child interactions. Such definitions and criteria should specify the salient features of each domain. It is not clear from the Grusec and Davidov presentation if the domains are to be defined by activities of the parents, types of parent–child interactions, goals of parents, and child outcomes, or some configuration of these three.

Bidirectionality, Reciprocity, and Development

Earlier I listed four problems in the global characterizations of parenting and indicated that the domain-specific propositions put forth by Grusec and Davidov mainly serve as a corrective to one problem, that of portraying parenting as global. Through their five domains and their emphasis on socialization, Grusec and Davidov raise questions pertaining to the other problems I listed—which revolve around processes of development and the nature of interactions between parents and children. Given the emphasis Grusec and Davidov place on the idea of bidirectionality in parent–child interactions and on domains that include reciprocity (parents and children interacting as equal status participants and accommodating one another's wishes), guided learning (pertaining to cognitive, social, and emotional understandings), and group participation (bringing in cultural practices and experiences beyond the family), it is surprising that they restrict themselves to socialization as children's outcomes.

Grusec and Davidov (p. 691) define socialization as "the process by which children acquire the social, emotional, and cognitive skills needed to function in the social community." Most of the child outcomes they associate with socialization are couched in terms that involve the child fitting into the existing system. They emphasize compliance with parental requests, acceptance of and obedience to cultural rules, internalization of values (taking on the parental/societal values as their own), adjustment to the group, and acceptance of routines, rituals, social customs, and cultural practices. Interestingly, it is the control domain that is seen to produce outcomes in children's moral and principled behaviors (in contrast with reciprocal relations, which in my view are more likely sources of moral development).

The concept of development should be brought into these analyses. Whereas domains such as protection and control lend themselves to socialization, the other domains lend themselves to a developmental analysis that includes the formation of thought—social and nonsocial—that allows for children to think about their environment, evaluate parental directives and actions, and attempt to influence changes in their social relationships. In this sense, development is not solely compliance and accommodation. Research has demonstrated that moral acquisition does not solely involve an internalization of values from parents and others, but more importantly the construction of concepts of welfare, fairness, and rights pertaining to how people ought to relate with each other (for reviews, see Smetana, 2006; Turiel, 1998). Outgrowths of moral concepts are cooperation and adaptation to the group, as well as thinking about how people should interact and about social arrangements and cultural practices that do and do not meet moral criteria.

A developmental perspective is based on a "strong" view of bidirectionality. The weak view (Kochanska, 1997) is that parents (or other adults) take into account and modify their behaviors in accord with characteristics of the child (e.g., a child's temperament, levels of fear or anxiety, preferences or inclinations). The strong view is that relationships are reciprocal in that there is give and take, communication, even argumentation, mutual influences, and engagement with each others' perspectives (Overton, in press; Piaget, 1970; Selman, 1980; Turiel, in press; Werner, 1957). This is not to say that reciprocal processes involve people of equal status or that the influences are balanced (Piaget, 1932). There are bidirectional processes in parent–child relationships that are not equal but do involve communication and some give and take. With age the imbalances change to the point that during adolescence there is more argumentation, conflict, and mutual influences. This is a realm in which social interactions outside the family, such as among peers, are likely to have an influence on interactions within the family. Indeed, Piaget (1932) maintained that in late childhood and adolescence, peer relations, because of greater equality, were of more influence in the development of moral judgments than adult–child relationships. If this were correct, then such changes in children's moral judgment would likely be applied by them to family interactions.

For adequate analyses of bidirectionality in social interactions and development, it is necessary to take into account children's domains of thought (Nucci, 2001; Smetana, 2006; Turiel, 1983, 2002), as well as domains of social interactions. As Grusec and Davidov state, the work on social domains focuses on different types of social judgment and reasoning and on how children acquire "a differentiated understanding of different kinds of social issues." However, in developmental analyses it is important to connect domains of social thought with social interactions. A number of observational studies demonstrate that the domains of social thinking are associated with different types of social interactions (Killen & Smetana, 1999; Nucci & Nucci, 1982a, 1982b; Nucci & Turiel, 1978; Nucci, Turiel, & Encarnacion-Gawrych, 1983; Tisak, Nucci, & Jankowski, 1996). Parent–child interactions contribute to the formation of domains of social thought (Nucci & Weber, 1995; Smetana, 1989). A developmental perspective that seriously accounts for bidirectionality needs to examine the intersection of the child's domains of thought and social intersections, including ones involving parents and children.

Social Harmony and Social Conflict

The emphasis in socialization perspectives on compliance, accommodation, adaptation, and internalization of values serves to skew characterizations of social interactions in the direction of social harmony and cooperation. This is because the developing child is seen as increasingly complying and accommodating to the group. Insofar as conflicts exist, they are due to a child's inadequate socialization so that he or she has not learned control of impulses nor come to adapt to society. From the socialization perspective conflicts might also occur between groups (societies, subcultures, cultures) that maintain different values and practices.

Although social harmony is a fact of social life, I believe that it is safe to say that conflict is also a fact of social life (Shantz & Hartup, 1992). Social conflicts are not solely a consequence of inadequate socialization nor only manifested by deviants who engage in antisocial behaviors. Social conflicts occur between groups, but they are also prevalent in interactions among members of a group or culture. Intragroup conflicts occur in the work place, political establishments, religions, and friendships (Turiel, 2002). Such conflicts are over many issues pertaining to personal decisions, societal directions, and cultural practices.

Conflicts also frequently pertain to the main topic of Grusec and Davidov's analysis—the family. Family disputes, disagreements, and conflicts occur between spouses (Gottman, Driver, Yoshimoto, & Rushe, 2002) and over how to relate to and raise children (Cowan & Cowan, 1992). Conflicts among adults occur in many cultures—especially patriarchal cultures—over societal arrangements and cultural practices that foster inequalities between males and females (Nussbaum, 1999; Okin, 1989). Research in several non-Western cultures shows that women judge as unfair practices that embed inequalities favoring males (Conry-Murray, 2009; Neff, 2001; Turiel & Wainryb, 2000; Wainryb & Turiel, 1994). Research also shows that women placed in subordinate positions in social hierarchies oppose, resist, and even engage in overt and covert acts of subversion to circumvent unjust practices (Abu-Lughod, 1993; Wikan, 1996). These are examples of how adults oppose in-group norms and practices.

Conflicts and acts of opposition that do not stem from incomplete or inadequate accommodation to parents occur within the family during childhood and adolescence. Children do develop positive social and moral orientations to others, act cooperatively, and adhere to social customs. Children's positive orientations coexist with acts of opposition, conflicts, and defiance of prohibitions established by parents (Dunn, 1987, 1988; Dunn, Brown, & Maguire, 1995). Conflicts and acts of opposition involve efforts on the part of children to evaluate directives from adults (Laupa, Turiel, & Cowan, 1995).

Opposition, resistance, and conflicts do not dissipate with age. In many ways they become more frequent, intense, and differentiated in adolescence. There is a vast literature on conflicts in adolescence, of course. Research relevant to social opposition and resistance include investigations of adolescents' disclosure and nondisclosure of activities to their parents (Kerr & Stattin, 2000; Smetana, Metzger, Gettman, & Campione-Barr, 2006; Stattin & Kerr, 2000) and adolescents' judgments about deception of parents (Perkins & Turiel, 2007). The research has documented that adolescents withhold information from parents in order to spare parents' feelings and to attain a sense of autonomy. Adolescents' decisions regarding disclosure of activities vary by domain (e.g., they are more likely to not disclose with regard to personal decisions than moral or prudential matters). Adolescents judge deception of parents as acceptable in realms they consider outside of legitimate parental authority (e.g., certain personal choices) but not in realms they consider parents to have legitimate authority to control (e.g., prudential matters like safety).

Concluding Remarks

I mention a few studies with children and adolescents that demonstrate that there is a coexistence of harmony and cooperation, on the one hand, and opposition and conflict, on the other hand. When children or adolescents "push back" in the sense that they oppose, criticize, or try to get around parental directives it is not always because they have not been adequately socialized. Some of the time, it is because they perceive some matters differently from their parents in ways that might be carried into productive choices when they grow into adulthood (Turiel & Perkins, 2004). This suggests that parent–child relationships are multifaceted, as are processes of thought, emotions, and development. The domain-specific propositions put forth by Grusec and Davidov provide a useful means for capturing these complexities in research on parent–child interactions. However, there is need for more precise definitions of the classification system and criteria for each domain in order to implement the ideas in research. I also believe that there is sufficient evidence in developmental research for the proposition that parent–child relationships involve the type of bidirectionality that requires taking into account the perspectives of children and how they interpret parental directives and actions. A bidirectional, reciprocal process of social interactions means that the development of thought, emotions, and actions need to be part of research and theory on parenting.

References

Abu-Lughod, L. (1993). *Writing women's worlds: Bedouin stories*. Berkeley: University of California Press.

Bandura, A., & Walters, R. (1963). *Social learning and personality development*. New York: Holt, Rinehart & Winston.

Bornstein, M. H. (2006). Parenting science and practice. In W. Damon & R. M. Lerner (Series Eds.) & I. E. Sigel & K. A. Renninger (Vol. Eds.), *Handbook of child psychology: Vol. 4. Child psychology in practice* (6th ed., pp. 893–949). New York: Wiley.

Chomsky, N. (1979). *Language and responsibility*. New York: Pantheon Books.

Conry-Murray, C. (2009). Adolescent and adult reasoning about gender roles and fairness in Benin, West Africa. *Cognitive Development, 24*, 207–219.

Cowan, P. A., & Cowan, C. P. (1992). *When partners become parents: The big life changes for couples*. New York: Basic Books.

Cowan, P. A., Powell, D., & Cowan, C. P. (1998). Parenting interventions: A family systems perspective. In W. Damon (Series Ed.) & I. E. Sigel & K. A. Renninger (Vol. Eds.), *Handbook of child psychology: Vol. 4. Child psychology in practice* (5th ed., pp. 3–72). New York: Wiley.

Crouter, A. C., & McHale, S. M. (2005). Work, family, and children's time: Implications for youth. In S. M. Bianchi, L. M. Casper & R. Berkowitz King (Eds.), *Work, family, and well-being* (pp. 47–64). New York: Routledge.

Dunn, J. (1987). The beginnings of moral understanding: Development in the second year. In J. Kagan & S. Lamb (Eds.), *The emergence of morality in young children* (pp. 91–112). Chicago: University of Chicago Press.

Dunn, J. (1988). *The beginnings of social understanding*. Cambridge, MA: Harvard University Press.

Dunn, J., Brown, J. R., & Maguire, M. (1995). The development of children's moral sensibility: Individual differences and emotion understanding. *Developmental Psychology, 31*, 649–659.

Freud, S. (1960). *The ego and the id*. New York: Norton. (Original work published 1923)

Gottman, J. M., Driver, J., Yoshimoto, D., & Rushe, R. (2002). Approaches to the study of power in violent and nonviolent marriages, and in gay male and lesbian cohabiting relationships. In P. Noller & J. A. Feney (Eds.), *Understanding marriage: Developments in the study of couple interactions. Advances in personal relationships* (pp. 323–347). Cambridge, UK: Cambridge University Press.

Grusec, J. E., & Davidov, M. (2010). Integrating different perspectives on socialization theory and research: A domain-specific approach. *Child Development, 81*, 687–709.

Kerr, M., & Stattin, H. (2000). What parents know, how they know it, and several forms of adolescent adjustment: Further support for a reinterpretation of monitoring. *Developmental Psychology, 36*, 366–380.

Killen, M. (1989). Context, conflict, and coordination in social development. In L. T. Winegar (Ed.), *Social interaction and the development of children's understanding* (pp. 119–146). Norwood, NJ: Ablex.

Killen, M., & Smetana, J. G. (1999). Social interactions in preschool classrooms and the development of young children's conceptions of the personal. *Child Development, 70*, 486–501.

Kochanska, G. (1997). Multiple pathways to conscience for children with different temperaments: From toddlerhood to age 5. *Developmental Psychology, 33*, 228–240.

Laupa, M., Turiel, E., & Cowan, P. A. (1995). Obedience to authority in children and adults. In M. Killen & D. Hart (Eds.), *Morality in everyday life: Developmental perspectives* (pp. 131–165). Cambridge, UK: Cambridge University Press.

Neff, K. D. (2001). Judgments of personal autonomy and interpersonal responsibility in the context of Indian spousal relationships: An examination of young people's reasoning in Mysore, India. *British Journal of Developmental Psychology, 19*, 233–257.

Nucci, L. P. (2001). *Education in the moral domain*. Cambridge, UK: Cambridge University Press.

Nucci, L. P., & Nucci, M. S. (1982a). Children's responses to moral and social conventional transgressions in free-play settings. *Child Development, 53*, 1337–1342.

Nucci, L. P., & Nucci, M. S. (1982b). Children's social interactions in the context of moral and conventional transgressions. *Child Development, 53*, 403–412.

Nucci, L. P., & Turiel, E. (1978). Social interactions and the development of social concepts in preschool children. *Child Development, 49*, 400–407.

Nucci, L. P., Turiel, E., & Encarnacion-Gawrych, G. (1983). Children's social interactions and social concepts: Analyses of morality and convention in the Virgin Islands. *Journal of Cross-Cultural Psychology, 14*, 469–487.

Nucci, L. P., & Weber, E. (1995). Social interactions in the home and the development of young children's conceptions of the personal. *Child Development, 66*, 1438–1452.

Nussbaum, M. C. (1999). *Sex and social justice*. New York: Oxford University Press.

Okin, S. M. (1989). *Justice, gender, and the family*. New York: Basic Books.

Overton, W. F. (in press). Life-span development: Concepts and issues. In W. F. Overton (Ed.), *Handbook of life-span development*. New York: Wiley.

Perkins, S. A., & Turiel, E. (2007). To lie or not to lie: To whom and under what circumstances. *Child Development, 78*, 609–621.

Piaget, J. (1932). *The moral judgment of the child.* London: Routledge & Kegan Paul.

Piaget, J. (1970). *Psychology and epistemology.* New York: Viking Press.

Ross, L., & Nisbett, R. M. (1991). *The person and the situation: Perspectives on social psychology.* Philadelphia: Temple University Press.

Rubin, K. H., Bukowski, W. M., & Parker, J. G. (2006) Peer interactions, relationships, and groups. In W. Damon & R. M. Lerner (Series Eds.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (6th ed., pp. 571–645). New York: Wiley.

Selman, R. L. (1980). *The growth of interpersonal understanding: Developmental and clinical analyses.* New York: Academic Press.

Shantz C. U., & Hartup W. W. (Eds.). (1992). *Conflict in child and adolescent development.* Cambridge, UK: Cambridge University Press.

Sigel, I. E., & McGillicuddy-DeLisi, A. V. (2002). Parental beliefs and cognitions: The dynamic belief systems model. In M. H. Bornstein (Ed.), *Handbook of parenting: Vol. 3. Status and social conditions of parenting* (2nd ed., pp. 458–508). Mahwah, NJ: Erlbaum.

Smetana, J. G. (1989). Toddlers' social interactions in the context of moral and conventional transgressions in the home. *Developmental Psychology, 25,* 499–508.

Smetana, J. G. (2006). Social domain theory: Consistencies and variations in children's moral and social judgments. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 119–153). Mahwah, NJ: Erlbaum.

Smetana, J. G., Metzger, A., Gettman, D. C., & Campione-Barr, N. (2006). Disclosure and secrecy in parent-adolescent relationships. *Child Development, 77,* 201–217.

Smetana, J. G., & Turiel, E. (2003). Moral development during adolescence. In G. R. Adams & M. D. Berzonsky (Eds.), *Blackwell handbook of adolescence* (pp. 247–268). Malden, MA: Blackwell.

Stattin, H., & Kerr, M. (2000). Parental monitoring: A reinterpretation. *Child Development, 71,* 1072–1085.

Tisak, M. S., Nucci, L. P., & Jankowski, A. M. (1996). Preschool children's social interactions involving moral and prudential transgressions: An observational study. *Journal of Early Education & Development, 7,* 137–148.

Turiel, E. (1983). *The development of social knowledge: Morality and convention.* Cambridge, UK: Cambridge University Press.

Turiel, E. (1998). The development of morality. In W. Damon (Series Ed.) & N. Eisenberg (Vol. Ed.) *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (5th ed., pp. 863–932). New York: Wiley.

Turiel, E. (2002). *The culture of morality: Social development, context, and conflict.* Cambridge, UK: Cambridge University Press.

Turiel, E. (2005). The many faces of parenting. In J. G. Smetana (Ed.), *Changing boundaries of parental authority during adolescence. New directions in child and adolescent development* (pp. 79–88). San Francisco: Jossey-Bass.

Turiel, E. (2008). Social decisions, social interactions, and the coordination of diverse judgments. In U. Mueller, J. I. Carpendale, N. Budwig, & B. Sokol (Eds.), *Social life and social knowledge: Toward a process account of development* (pp. 255–276). Mahwah, NJ: Erlbaum.

Turiel, E. (in press). The development of morality: Reasoning, emotions, and resistance. In W. F. Overton (Ed.), *Handbook of life-span development.* New York: Wiley.

Turiel, E., & Perkins, S. A. (2004). Flexibilities of mind: Conflict and culture. *Human Development, 47,* 158–178.

Turiel, E., & Wainryb, C. (2000). Social life in cultures: Judgments, conflicts, and subversion. *Child Development, 71,* 250–256.

Wainryb, C., & Turiel, E. (1994). Dominance, subordination, and concepts of personal entitlements in cultural contexts. *Child Development, 65,* 1701–1722.

Werner, H. (1957). *Comparative psychology of mental development.* New York: International Universities Press.

Wikan, U. (1996). *Tomorrow, God willing: Self-made destinies in Cairo.* Chicago: University of Chicago Press.

Child Development, May/June 2010, Volume 81, Number 3, Pages 727–736

Connecting Cues: Overlapping Regularities Support Cue Discovery in Infancy

Sarah D. Sahni, Mark S. Seidenberg, and Jenny R. Saffran
University of Wisconsin–Madison

The present work examined the discovery of linguistic cues during a word segmentation task. Whereas previous studies have focused on sensitivity to individual cues, this study addresses how individual cues may be used to discover additional, correlated cues. Twenty-four 9-month-old infants were familiarized with a speech stream in which syllable-level transitional probabilities and an overlapping novel cue served as cues to word boundaries. Infants' behavior at test indicated that they were able to discover the novel cue. Additional experiments showed that infants did not have a preexisting preference for specific test items and that transitional probability information was necessary to acquire the novel cue. Results suggest one way learners can discover relevant linguistic structure amid the multiple overlapping properties of natural language.

Natural languages exhibit structure at multiple levels in parallel (e.g., phonological, lexical, morphological, syntactic, and discourse). For the adult listener, this complexity creates temporary ambiguities that must be resolved for speech to be understood. Individual bits of information are imprecise, such as the meaning of words like *bow*, *colon*, *saw*, and *wave*. Such ambiguities are resolved via a constraint satisfaction process that exploits correlations among different types of information (MacDonald, Pearlmutter, & Seidenberg, 1994). While individual cues are often unreliable, combinations of cues are not. The principal characteristic of the constraint satisfaction process is that it allows learners to utilize the correlations of cues. For example, the word *saw* has several meanings (related to seeing, cutting, a tool for cutting, etc.) and is thus highly ambiguous in isolation. Embedded within an utterance such as ''I saw you,'' ''I'' restricts its interpretation to verbs. The object ''you'' further restricts the interpretation to the ''seeing'' meaning, since a person is more likely to be seen than sawed, though the result might differ if the context were a magic show. Similarly, hearing *saw* in a hardware store suggests the noun interpretation of *saw* as a tool. Rapid online comprehension is possible because of our ability to exploit constraints between different types of information (Seidenberg & MacDonald, 1999).

While studies of adult language have investigated how constraints are combined to resolve ambiguities, studies of language acquisition have examined how children use statistical cues to learn their native language. These are complementary issues, the ''constraints'' that are relevant to adult listeners are the ''cues'' by which the child acquires language (Seidenberg, 1997). Seminal work by Saffran, Aslin, and Newport (1996) showed that 8-month-olds are sensitive to the transitional probability (TP) between two syllables (the frequency of the two syllables divided by the frequency of the first syllable) when listening to a fluent stream of speech. This study, and the large body of work that has followed it, suggests that infants are sensitive to statistical regularities that exist in natural languages and can use them to learn aspects of their native language.

Experiments investigating language acquisition via statistical learning have typically focused on infants' abilities to use one statistical cue at a time

---

We thank Allison Dahlke, Catherine Moore, and Diana Dorovany for help with stimulus construction and subject running; Katherine Kortenkamp for consultation on statistical analyses; an anonymous referee for the suggestion of the chi-square analysis in Experiment 1; and Jessica Hay, Stephen Laniel, Rachel Sussman, Erik Thiessen, and four anonymous referees for helpful comments on an earlier version of this manuscript. This work was supported by NIH Grant F31DC008737 to S.D.S., NIH Grant RO1HD37466 to J.R.S., and NICHD Core Grant P30HD03352 to the Waisman Center.

Correspondence concerning this article should be addressed to Sarah D. Sahni, Department of Psychology, University of Wisconsin–Madison, 1202 W. Johnson St., Madison, WI 53706. Electronic mail may be sent to sdsahni@gmail.com.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8103-0006

(e.g., Curtin, Mintz, & Christiansen, 2005; Saffran et al., 1996). However, natural speech is complex, containing overlapping regularities at multiple levels. For the language learner, this presents a difficult problem: How are these cues discovered? There are many ways speech can be analyzed; how does the child determine which aspects of the input are relevant? Moreover, the fact that any given bit of information may contribute to multiple levels of analysis (e.g., */b/* is the first sound in the word *baby*, the beginning of the first syllable, receives primary stress making it louder and longer, the transition point between the words *the baby*, etc.) creates a difficult learning problem. The complexity of this learning problem is sometimes thought to limit severely the explanatory role played by statistical learning in language acquisition (Yang, 2004).

Alternatively, the constraint satisfaction approach suggests that the complexity of natural language provides a rich system for learning mechanisms to exploit. Linguistic regularities reinforce each other across levels, allowing statistical learning mechanisms to capitalize on multiple cues and redundancies. For example, lexical stress patterns are found in numerous languages of the world. These patterns consist of a specific ordering of strong and weak syllables that occur frequently and can help identify word boundaries or classify groups of words. In English, many words have a trochaic, or strong–weak, stress pattern, as in the words *BAby* and *MOmmy* (Cutler & Carter, 1987). In other languages, it is more common for words to have an iambic (weak–strong) stress pattern, as in the word *guiTAR*. And in some languages stress cannot be used to group syllables or identify word boundaries at all. Because these cues vary from language to language, they must be learned. How then does the infant discover that stress patterns are informative? Strong regularities like lexical stress overlap with other regularities at multiple levels, highlighting and reinforcing their utility. The acoustic regularities (i.e., higher pitch, longer duration, and increased volume) of stressed syllables can draw attention to the beginning of trochaic words. Distributional cues, such as the overrepresentation of trochaic items in speech to English-learning children, ensure that young language learners have plenty of exposure to the pattern. Together these regularities can enhance the accessibility of the lexical stress pattern.

Psycholinguistic studies support the hypothesis that infants are sensitive to the conjunction of multiple probabilistic cues (Gerken, Wilson, & Lewis, 2005; Shi, Werker, & Morgan, 1999). Additionally, studies of infant categorization and conceptual development demonstrate that the natural environment provides infants with a multitude of correlated cues that they are able to exploit (e.g., Bhatt, Wilk, Hill, & Rovee-Collier, 2004; Madole, Oakes, & Cohen, 1993; Rakison, 2004; Younger, 1992; Younger & Cohen, 1986). Finally, connectionist models have shown that simple learning mechanisms that capitalize on structure within a complex system can exploit multiple correlated cues that exist in the infants' world. Computational models have demonstrated that problems such as finding word boundaries (Christiansen, Allen, & Seidenberg, 1998), generating properly inflected forms (Joanisse & Seidenberg, 1999; Mirkovic, MacDonald, & Seidenberg, 2005), and grouping common objects into categories (Rogers & McClelland, 2004) can be solved using multiple cues.

Despite this progress, it remains to be determined how language learners isolate and combine cues given the complexity of human language. In considering this problem, it is helpful to distinguish between two classes of potentially useful cues: language-general cues and language-specific cues. Measures of co-occurrence or predictability between syllables are language-general cues, in that they operate in similar ways across natural languages. For example, transitional probabilities are not specific to any given language (though the units over which these computations are performed are).

Other cues may or may not be useful in any given language and are thus language specific, and must be learned. For example, languages have different lexical stress patterns (iambic vs. trochaic), and in some languages stress patterns do not mark boundaries or help individual units cohere. By 9 months of age, infants typically show sensitivity to a range of language-specific cues (for a recent review, see Saffran & Sahni, in press).

In the domain of word segmentation, previous work suggests that younger infants tend to use language-general cues and later shift to language-specific cues (Thiessen & Saffran, 2003). In a segmentation study using a nonsense language, TPs, a language-general cue operating over novel syllable combinations was placed in conflict with the language-specific stress pattern of English. Six-and-a-half month-olds segmented the fluent speech using the language-general strategy of relying on TPs. In contrast, infants who were 2 months older used language-specific lexical stress patterns (also see Johnson & Jusczyk, 2001). This shift suggests that over time, infants become more sensitive to idiosyncratic cues, learning which regularities are relevant (and, presumably, which are not) for their native language. However, little is known about how this process unfolds.

How might infants discover these language-specific cues? One potential explanation is that language-general cues provide a basis for discovering overlapping or co-occurring language-specific cues. For example, in word segmentation, infants may use their sensitivity to TP cues, which is present early in life (Kirkham, Slemmer, & Johnson, 2002; Teinonen, Fellman, Naatanen, Aklu, & Huotilainen, 2009) to discover language-specific cues that are correlated with TPs.

The present work tested the hypothesis that infants can discover novel cues by exploiting redundancies between language-general and language-specific cues. Nine-month-old infants were exposed to a fluent speech stream that contained two overlapping cues to word boundaries: a language-general cue (TPs) and a language-specific cue (/t/-onsets). TPs are known to be salient to 9-month-old infants. The second cue was specific to the artificial language and therefore novel: Each word in the speech stream began with /t/. Experiment 1 was thus designed to test the hypothesis that infants can use the language-general TP cue to discover the overlapping language-specific /t/-onset cue. The /t/-initial syllables are only informative as a cue to word boundaries due to their overlap with the TP cue; the TP cue positions the /t/-initial syllable at the onset of each word. Consequently, the only way infants can extract this pattern is to use its overlap with the TP cue. We tested infants using items that were all novel relative to the exposure corpus but that varied in their use of the /t/-onset cue. The question of interest was whether infants would be sensitive to the presence of /t/-onsets in the test items. If so, this would provide evidence that infants can isolate individual cues by using redundancies in the speech stream.

Experiment 1

To examine whether infants can use a language-general segmentation cue to discover an overlapping novel language-specific cue, infants heard a fluent speech stream that contained two overlapping cues to word boundaries: (a) dips in TPs at word boundaries and (b) /t/-initial syllables at word onsets. To determine whether infants acquired the novel /t/-onset pattern, test items either adhered to the pattern (began with a /t/-syllable) or violated it (contained a medial /t/). Crucially, these items were previously unheard combinations of syllables from the speech stream (i.e., TP = 0). Therefore, TP information would not allow infants to distinguish between the two types of test items. Instead, successful discrimination hinged on discovery of the /t/-initial pattern present in the speech stream played during familiarization.

Method

Participants. Twenty-four 9.5-month-old monolingual English-learning infants (mean age = 9.5 - months, range = 9.0–10.0) participated in this experiment. All infants were born full-term and had fewer than four prior ear infections and no history of hearing or vision impairments. Data from an additional 8 infants were excluded due to fussiness (4) and parents stopping the experiment (4).

Stimuli. A fluent stream of speech was created from recordings of a female native English speaker who was blind to the structure of the artificial language. The language contained six bisyllabic words: *tohsigh, teemay, tiepu, tukee, tayla,* and *tafo.* A pseudosynthesis technique was used to create the speech stream, which allowed for use of naturally produced syllables while permitting control over coarticulation, duration, pitch, and volume of all syllables in the language. All three-syllable sequences that occurred in the language, both within and between word boundaries (e.g., *tohsigh-tee, sighteemay*), were recorded in a monotone, isochronous register. Medial syllables were spliced out of the three-syllable sequences and concatenated together with no silence between syllables. By using these medial syllables, coarticulation within each syllable and between every pair of syllables in the language was maintained. Syllables were edited prior to concatenation to have the same duration, pitch, and volume. The stream contained 40 repetitions (2 min 17 s) of each word in a pseudorandom order with no word appearing twice in succession (see the Appendix for a transcript of the complete familiarization language). Each within-word syllable pair had a TP of 1.0; between-word syllable pairs had a TP between .1 and .25 ($M = .20$). The speech stream thus contained two overlapping and completely redundant cues to word boundaries: dips in TPs and /t/-onsets.

Four novel test items were constructed from syllables in the artificial language. Two of these items began with /t/ (*tiemay, tohla*), and two contained a medial /t/ (*fota, keetu*). Test items were created from recordings of each bisyllabic item spoken in isolation. Duration, pitch, and volume were edited so that all test items were essentially equivalent.

Procedure. During familiarization, infants listened to the speech stream at a comfortable volume, played over speakers mounted on each sidewall while viewing an unrelated Baby Einstein video. An experimenter then entered the booth, covered the monitor that displayed the video, and placed headphones playing masking music on the caregiver. The test phase began with two practice trials (a recording of piano tones), designed to help the infants learn the contingency between their head-turns and the lights and sounds. The practice trials were followed by 12 test trials, three blocks of each of the four test items (*tiemay*, *tohla*, *fota*, and *keetu*). Infants' ability to discriminate the test items was assessed using the Headturn Preference Procedure (Kemler Nelson et al., 1995). The experimenter was seated outside the booth, observing the infants' head-turns on a closed circuit TV, and controlling the experiment via custom software. Lights were mounted on the center wall (directly facing the infant) and sidewalls. Each trial began with the center light blinking. Once the infant fixated on the light, it was extinguished and one of the sidelights began to blink. When the infant fixated on the blinking sidelight a sound was played from the speaker below the light. On each test trial, an item was repeated until the infant looked away for at least 2 s, or until the item had repeated 15 times. If the infant failed to fixate on the side light for at least 1 s during a test trial, the trial was excluded and an additional trial of that test item was automatically added after the third test block.

Results

We tested infants' ability to discriminate /*t*/-initial from /*t*/-medial test items over the three blocks of testing with a 2 (test item type: /*t*/-initial vs. /*t*/-medial) × 3 (test block: 1, 2, 3) repeated-measures analysis of variance (ANOVA; means shown in Table 1). The main effect of test item type (/*t*/-initial vs. /*t*/-medial) was not significant, $F(1, 23) = 0.975$, $p = .33$. The assumption of sphericity was violated for the Type × Block interaction term (Mauchly's $W = .687$) and so multivariate analyses were used to evaluate the significance of the interaction. With a large violation of sphericity (i.e., when Mauchly's $W < .7$), the statistical power of multivariate techniques tends to be greater than univariate techniques (Keppel, 1991; Mendoza, Toothaker, & Nicewander, 1974). There was a significant interaction between test item type and block, $F(2, 22) = 8.15$, $p = .002$. As shown in Figure 1, the significant interaction reflects a reversal in the direction of preference over the course of testing. The familiarity preference present in the first two test blocks shifts to a novelty preference in the third block. Block interactions and shifts in direction of preference have been previously observed elsewhere in the literature (Gerken et al., 2005) but are not often discussed.

Subsequent analyses focused on the first two blocks (eight test trials), as these looking times are more proximal to the familiarization phase and thus most likely to reflect learning from the fluent speech. A one-way (test item type: /*t*/-initial vs. /*t*/-medial) repeated-measures ANOVA revealed a significant difference in looking times to the two types of items, $F(1, 23) = 6.30$, $p = .02$ (see Figure 2). Infants looked longer to the /*t*/-initial test items, which adhered to the pattern presented during familiarization. Recall that the TPs between syllables in the test items were all zero. Thus, infants could not have discriminated /*t*/-initial from /*t*/-medial test items based on TP cues. These results suggest that infants learned the /*t*/-initial pattern and generalized it to include the novel test items.

Discussion

The results of Experiment 1 suggest that infants were able to exploit the /*t*/-initial pattern,

Table 1
Mean and Standard Errors for Looking Times

	Experiment 1		Experiment 2		Experiment 3	
	/*t*/-initial	/*t*/-medial	/*t*/-initial	/*t*/-medial	/*t*/-initial	/*t*/-medial
Block 1	8.69 (0.80)	7.54 (0.70)	8.44 (0.88)	8.75 (0.70)	8.30 (0.96)	7.75 (0.66)
Block 2	7.45 (0.71)	6.02 (0.75)	7.00 (0.57)	7.56 (0.70)	6.77 (0.56)	6.95 (0.68)
Block 3	4.58 (0.39)	5.94 (0.56)	6.79 (0.68)	6.32 (0.79)	7.18 (0.85)	6.77 (0.73)
Blocks 1 and 2 (averaged)	8.07 (0.58)	6.78 (0.59)	7.57 (0.60)	8.18 (0.53)	7.53 (0.58)	7.35 (0.55)
All trials (averaged)	6.92 (0.46)	6.53 (0.46)	7.40 (0.51)	7.52 (0.48)	7.31 (0.39)	7.09 (0.50)

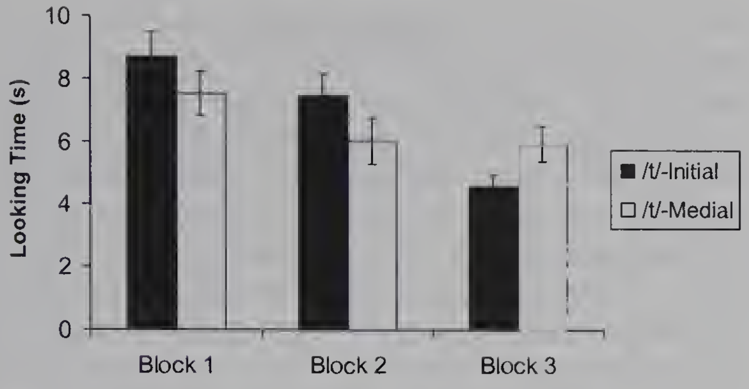


Figure 1. Experiment 1: Mean looking times to /*t*/-initial and /*t*/-medial test items for each test block.

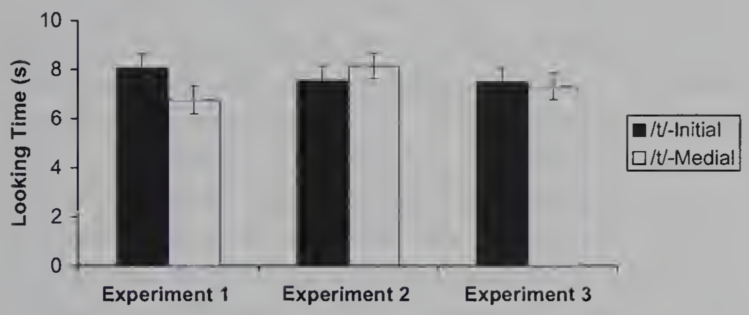


Figure 2. The average looking time for /*t*/-initial and /*t*/-medial items across test Blocks 1 and 2 for all three experiments.

successfully discriminating novel test items that followed the pattern from those that did not. This pattern was not immediately obvious in the input; the speech stream consisted of syllables beginning with /*t*/ alternating with syllables that began with other sounds. In order to discover that the /*t*/ segment signaled word onsets, infants presumably capitalized on the TP cues in the speech stream, which also provided cues to word boundaries. On this view, infants discovered the language-specific /*t*/-onset cue by capitalizing on the language-general TP cue.

One interesting feature of these data is that infants' looking behavior changed over the course of testing. Familiarity to novelty preference switches is not uncommon in infant behavioral studies, though the factors responsible for the shift may vary (Fantz, 1964; Hunter, Ames, & Koopman, 1983). The test items used in this study consisted of novel combinations of syllables from the familiarization language. This introduction of novel items at test forces infants to generalize beyond the training corpus, making it more likely that participants will show a familiarity preference at the outset of testing (e.g., Thiessen & Saffran, 2003). However, during the course of testing, infants received differential amounts of exposure to the test items. Figure 3 depicts the difference in looking times to /*t*/-initial and /*t*/-medial test items across the three test blocks. Initially, infants looked longer to /*t*/-initial items, thus receiving more exposure to them than to the /*t*/-medial items. This pattern of listening during testing may have led the infants to become bored with these items, moving them toward a novelty preference. To test this hypothesis, we examined individual participants' looking preferences across the testing session. Fifteen of 24 participants showed an initial familiarity preference that transitioned into a novelty preference, 5 showed an initial familiarity preference that remained a familiarity preference, and 4 showed an initial novelty preference that remained a novelty preference. A majority of participants showed the predominant pattern of increased exposure to the /*t*/-initial items in the first two blocks with a novelty preference in the third block. A chi-square test confirmed that this pattern of behavior would not be expected by chance ($\chi^2 = 20.3$, $df = 3$, $p = .0001$). This pattern of results suggests that while infants' initial test responses were linked to learning during the familiarization phase, the novelty preference in Block 3 may have reflected infants' experiences during testing.

The findings from Experiment 1 support the hypothesis that infants used low TPs at word boundaries to acquire the overlapping but novel /*t*/-onset pattern. However, there is an alternate explanation for these results: It is possible that infants' behavior reflected preexisting preferences for individual items. A counterbalanced language composed of items that all contain a medial /*t*/ would clarify this issue. However, previous work (Endress, Scholl, & Mehler, 2005) has shown that it may be easier to generalize from patterns that occur at the edges of sequences, as opposed to those

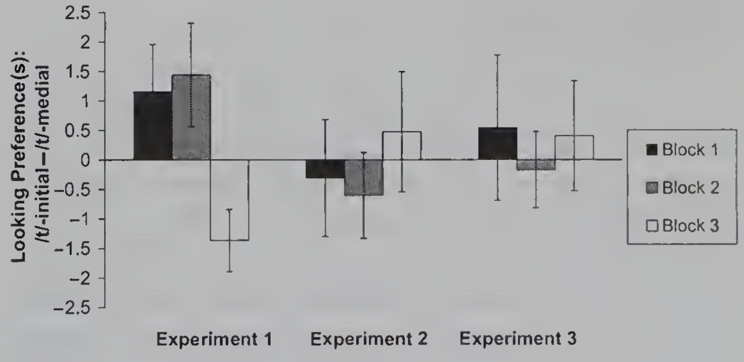


Figure 3. The difference between looking times to /*t*/-initial and /*t*/-medial test items, for each block for all three experiments.
Note. Positive values indicate a familiarity preference; negative values indicate a novelty preference.

occurring medially. Consequently, the two counterbalanced languages might not be equally learnable. A second experiment tested this possibility with a new group of infants, who participated only in the test phase of the experiment. If infants in Experiment 1 had an a priori preference for the /t/-initial test items, infants in Experiment 2 should show a similar pattern of behavior. If infants in Experiment 2 do not show a similar pattern, this would suggest that exposure to familiarization materials that contained the two overlapping cues was necessary to elicit the preference for /t/-initial test items.

Experiment 2

This experiment was designed to test the hypothesis that infants in Experiment 1 listened longer to the /t/-initial test items due to an a priori preference for these particular items. Infants in Experiment 2 were not exposed to the familiarization speech stream, participating only in the testing procedure used in Experiment 1.

Method

Participants. Twenty-four 9.5-month-old monolingual English-learning infants (mean age = 9.5 months, range = 9.0–9.9) participated. Data from an additional 5 infants were excluded from the analyses because of experimenter error (1), fussiness (2), failure to contribute at least two trials for each item (1), and mean looking time to one or both sides less than 3 s (1).

Stimuli. The test items were the same as those used in Experiment 1.

Procedure. There was no exposure phase. The testing procedure was identical to Experiment 1, with 2 practice trials followed by 12 test trials.

Results and Discussion

As in Experiment 1, a 2 (test item type: /t/-initial vs. /t/-medial) × 3 (test block: 1, 2, 3) repeated-measures ANOVA was conducted. There was no significant effect of test item type (/t/-initial vs. /t/-medial), $F(1, 23) = 0.081$, $p = .78$, nor was the interaction between block and test item type significant, $F(2, 46) = 0.35$, $p = .70$ (see Table 1 and Figure 3). These results indicate that infants in Experiment 2 did not discriminate between /t/-initial and /t/-medial test items.

We next conducted a 2 (test item type: /t/-initial vs. /t/-medial) × 2 (group: Experiment 1, Experiment 2) repeated-measures ANOVA over the data from the first two test blocks from Experiments 1 and 2. This analysis was intended to determine whether the behavior of infants differed reliably across the two experiments. The interaction between test item type and group was significant, $F(1, 46) = 6.19$, $p = .017$, suggesting that infants who heard the familiarization materials showed a different pattern of behavior at test than those who did not (see Figure 2).

This between-group analysis, coupled with the within-subject analysis showing no effect of test item type, indicates that infants in Experiment 2 did not have an a priori preference for the /t/-initial items relative to the /t/-medial items. We can therefore attribute infants' behavior in Experiment 1 to familiarization with the fluent speech stream. Nevertheless, it is still unclear which aspects of the familiarization stimuli elicited infants' successful discrimination between /t/-initial and /t/-medial test items. It is possible that, as originally hypothesized, low TPs at word boundaries anchor the alternating /t/ syllables, allowing infants to extract the /t/-onset cue and generalize it to the novel test items. Another possibility is that infants are extracting the extremely regular alternating /t/ syllable pattern (created because each of the bisyllabic words begins with a /t/) and uniformly mapping the /t/ syllable to word onsets. According to this alternative hypothesis, infants could capitalize on a systematic pattern (/t/-onsets) without the aid of another cue. On this account, they detect the regular alternation and map it onto onsets, potentially because onsets are privileged perceptually and/or lexically (e.g., Brent & Cartwright, 1996; Jusczyk, Jusczyk, Kennedy, Schomberg, & Koenig, 1995; Magnuson, Dixon, Tanenhaus, & Aslin, 2007; Marslen-Wilson & Zwitserlood, 1989).

To explore this hypothesis, we designed a new speech stream to determine whether infants could extract the novel /t/-onset pattern without the aid of another cue for bootstrapping. The resulting speech stream did not have low TPs at word boundaries but still contained the /t/-onset pattern. If infants can extract the /t/-onset pattern without an overlapping cue, infants in Experiment 3 should show a significant difference in looking time to /t/-initial items compared to /t/-medial items. However, if TPs played a critical role in the discovery of the novel pattern via bootstrapping, infants should not show a significant difference in looking time to /t/-initial items compared to /t/-medial items in the absence of TP cues.

Experiment 3

This experiment was designed to determine whether the overlapping cue from Experiment 1, low TPs at word boundaries, was necessary for infants to acquire the novel /t/-onset pattern. Infants were exposed to a new fluent speech stream that did not have low TPs at word boundaries but still contained the novel /t/-onset pattern. The procedure and test items were identical to Experiment 1.

Method

Participants. Twenty-four 9.5-month-old monolingual English-learning infants (mean age = 9.5 months, range = 9.1–10.0) participated in this experiment. Data from an additional 17 infants were excluded from the analyses because of parental interference (5), sleepiness (1), external noise (1), fussiness (7), and failure to contribute at least one trial for each item type in every block (3).

Stimuli. A fluent speech stream was created using the procedure and words (*tohsigh*, *teemay*, *tiepu*, *tukee*, *tayla*, and *tafo*) from Experiment 1. Again, all syllables in the language were measured and edited such that duration, pitch, and volume were equivalent for all syllables. Unlike Experiment 1, in which words were repeated in a random order, the words in this fluent speech stream were repeated in exactly the same order 40 times (*teemaytieputukeetafotaylatohsigh**teemaytieputukeetafotaylatohsigh**. . .*; see Curtin et al., 2005, for another example of this method). This method generated a 2 min 7 s stream in which every pair of syllables had a TP of 1.0 and every other syllable in the language began with a /t/. The test items were identical to those in Experiments 1 and 2 (*tiemay*, *tohla*, *fota*, and *keetu*).

Procedure. The experimental procedure was the same as Experiment 1.

Results and Discussion

As in the previous experiments, we first ran a 2 (test item type: /t/-initial vs. /t/-medial) × 3 (test block: 1, 2, 3) repeated-measures ANOVA. There was no significant effect of test item type, $F(1, 23) = 0.07$, $p = .80$, and the interaction with block was not significant, $F(1, 23) = 0.27$, $p = .61$ (see Table 1 and Figure 3). This pattern of results indicates that infants did not differentiate between /t/-initial and /t/-medial test items in the absence of the TP cue. We next ran a 2 (test item type: /t/-initial vs. /t/-medial) × 2 (group: Experiment 1, Experiment 3) repeated-measures ANOVA contrasting the data from the first two test blocks of Experiments 1 and 3. This Test Item Type × Group interaction did not reach significance, $F(1, 46) = 2.87$, $p = .20$ (see Figure 2).

Though the between-group analysis is inconclusive, infants' failure to discriminate the test items in Experiment 3 is consistent with the hypothesis that infants were influenced by the presence of TP cues in Experiment 1. Without the low TPs at word boundaries to anchor the alternating /t/ syllables to segment onsets, infants seem unable to extract the novel pattern.

General Discussion

The results of these experiments indicate that infants are able to use a language-general regularity (dips in TPs at word boundaries) to discover a second, language-specific regularity (/t/-onsets). Moreover, infants generalized this newly learned cue to novel items, as demonstrated by their test performance. These findings suggest one possible class of solutions to the learning problem described earlier: How do infants discover relevant linguistic cues when there are many way to analyze speech and a single bit of information can be informative at multiple levels? Just as adults use correlations among cues to resolve ambiguities when using language, infants are able to use such correlations to acquire language, as suggested by the constraint satisfaction approach. Thus, infants' learning capacities, such as the ability to encode correlations across different types of information (Rose & Ruff, 1987), seem well matched to properties of natural language. What seems initially to be an insurmountable barrier to learning—the fact that elements of language contribute to multiple levels of structure simultaneously—actually helps solve the language acquisition problem (see Hirsh-Pasek, Golinkoff, Hennon, & Maguire, 2004, for a similar example in word learning).

Research over the past decade has shown that infants are sensitive to many different patterns that can be informative for language learning. It remains unclear, however, how infants find individual cues and combine different types of information to understand the complex structure of their language. Behavioral research on multiple cue usage in this domain has typically taken the form of cue-conflict studies, examining relative reliance on different types of information over time (e.g., Johnson & Jusczyk, 2001; Mattys, Jusczyk, Luce, & Morgan, 1999;

Thiessen & Saffran, 2003). Although studies using this approach have been quite revealing, they cannot address how infants may capitalize on the redundancies in natural speech or how cues are discovered. More recent research has focused on the use of multiple probabilistic patterns to categorize lexical items (Gerken et al., 2005; Gómez & Lakusta, 2004; Shi et al., 1999). Results from the present work demonstrate that the complexity of natural speech does not necessarily hinder language acquisition but in fact may facilitate learning. Paradoxically, complexity may help learning—as long as the complexity is consistent with the structure to be acquired (Morgan, Meier, & Newport, 1987, 1989).

The process observed in the present work is reminiscent of bootstrapping: Partial information about one element of language provides evidence about another element, which in turn provides further evidence for the first element (Gleitman & Wanner, 1982). On this view, the infant begins to pick up on TPs, facilitating discovery of the */t/*-onset cue, which in turn may further the consolidation of the TP cue. Both the TPs and the */t/*-onset cue provided evidence about the boundaries between words in the fluent speech stream. The */t/* cue is different, insofar as its discovery depended on some prior learning about TPs. We are not claiming that TPs or any specific regularity is necessary for this process to operate. Rather, any cues that infants can use to extract linguistic structure, and that overlap with other discoverable patterns, should be available for use in this fashion. This is an area where computational models of bootstrapping mechanisms would be informative. Existing models have typically built in different types of regularities, focusing on how combinations of given regularities can yield better learning outcomes than individual regularities (Christiansen et al., 1998; Shi, Morgan, & Allopenna, 1998). Models of how the cues themselves are identified, and the dependencies between them in learning, would be very timely.

This work shows that infants can use the overlapping nature of speech to isolate cues. Language-general regularities, such as TPs, may support the discovery of language-specific cues. Critically, the present research also expands the scope of statistical learning mechanisms. Not only can infants use such mechanisms to exploit the structure of their language, but they can also use statistical learning to discover the structure of their native language.

References

Bhatt, R. S., Wilk, A., Hill, D., & Rovee-Collier, C. (2004). Correlated attributes and categorization in the first half-year of life. *Developmental Psychobiology, 44*, 103–115.

Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition, 6*, 93–125.

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes, 13*, 221–268.

Curtin, S., Mintz, T. H., & Christiansen, M. H. (2005). Stress changes the representational landscape: Evidence from word segmentation. *Cognition, 96*, 233–262.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language, 2*, 133–142.

Endress, A. D., Scholl, B. J., & Mehler, J. (2005). The role of salience in the extraction of algebraic rules. *Journal of Experimental Psychology: General, 134*, 406–419.

Fantz, R. L. (1964). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science, 146*, 668–670.

Gerken, L., Wilson, R., & Lewis, W. (2005). Infants can use distributional cues to form syntactic categories. *Journal of Child Language, 32*, 249–268.

Gleitman, L. R., & Wanner, E. (Eds.). (1982). *Language acquisition: The state of the art*. New York: Cambridge University Press.

Gómez, R. L., & Lakusta, L. (2004). A first step in form-based category abstraction by 12-month-old infants. *Developmental Science, 7*, 567–580.

Hirsh-Pasek, K., Golinkoff, R. M., Hennon, E., & Maguire, M. J. (2004). Hybrid theories at the frontier of developmental psychology: The emergentist coalition model of learning as a case in point. In D. G. Hall & S. R. Waxman (Eds.), *Weaving a lexicon* (pp. 173–204). Cambridge, MA: MIT Press.

Hunter, M. A., Ames, E. W., & Koopman, R. (1983). Effects of stimulus complexity and familiarization time on infant preferences for novel and familiar stimuli. *Developmental Psychology, 19*, 338–352.

Joanisse, M. F., & Seidenberg, M. S. (1999). Impairments in verb morphology following brain injury: A connectionist model. *Proceedings of the National Academy of Sciences of the United States of America, 96*, 7592–7597.

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language, 44*, 548–567.

Jusczyk, P. W., Jusczyk, A. M., Kennedy, L. J., Schomberg, T., & Koenig, N. (1995). Young infants' retention of information about bisyllabic utterances. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 822–836.

Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior & Development, 18*, 111–116.

Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition, 83*, B35–B42.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*, 676–703.

Madole, K. L., Oakes, L. M., & Cohen, L. B. (1993). Developmental changes in infants' attention to function and form-function correlations. *Cognitive Development, 8*, 189–209.

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science, 31*, 133–156.

Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 576–585.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology, 38*, 465–494.

Mendoza, J. L., Toothaker, L. E., & Nicewander, W. A. (1974). A Monte Carlo comparison of the univariate and multivariate methods for the two-way repeated measure design. *Multivariate Behavioral Research, 9*, 165–178.

Mirkovic, J., MacDonald, M. C., & Seidenberg, M. S. (2005). Where does gender come from? Evidence from a complex inflectional system. *Language and Cognitive Processes, 20*, 139–167.

Morgan, J. L., Meier, R. P., & Newport, E. L. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology, 19*, 498–550.

Morgan, J. L., Meier, R. P., & Newport, E. L. (1989). Facilitating the acquisition of syntax with cross-sentential cues to phrase structure. *Journal of Memory and Language, 28*, 360–374.

Rakison, D. H. (2004). Infants' sensitivity to correlations between static and dynamic features in a category context. *Journal of Experimental Child Psychology, 89*, 1–30.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach.* Cambridge, MA: MIT Press.

Rose, S. A., & Ruff, H. A. (1987). Cross-modal abilities in human infants. In J. D. Osofsky (Ed.), *Handbook of infant development* (2nd ed., pp. 318–362). Oxford, UK: Wiley.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.

Saffran, J. R., & Sahni, S. D. (in press). Learning the sounds of language. In M. Joanisse, M. Spivey, & K. McCrae (Eds.), *Cambridge handbook of psycholinguistics.* Cambridge, UK: Cambridge University Press.

Seidenberg, M. S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science, 275*, 1599–1603.

Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science, 23*, 569–588.

Shi, R., Morgan, J. L., & Allopenna, P. (1998). Phonological and acoustic bases for earliest grammatical category assignment: A cross-linguistic perspective. *Journal of Child Language, 25*, 169–201.

Shi, R., Werker, J. F., & Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition, 72*, B11–B21.

Teinonen, T., Fellman, V., Naatanen, R., Aklu, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience, 10*, 21.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology, 39*, 706–716.

Yang, C. D. (2004). Universal grammar, statistics or both? *Trends in Cognitive Sciences, 8*, 451–456.

Younger, B. A. (1992). Developmental change in infant categorization: The perception of correlations among facial features. *Child Development, 63*, 1526–1535.

Younger, B. A., & Cohen, L. B. (1986). Developmental change in infants' perception of correlations among attributes. *Child Development, 57*, 803–815.

Appendix

Transcript of Speech Stream From Experiment 1

teemaytohsighteemaytukeetafotieputaylatohsightayla
tukeeteemaytieputaylateemaytukeeteemaytohsightu
keetafoteemaytukeetohsighteemaytohsighteemaytoh
sightafoteemaytafotohsightukeetieputukeetieputafo
teemaytohsighteemaytukeetafotieputaylatohsightayla
tukeeteemaytieputaylateemaytukeeteemaytohsi tee
maytohsighteemaytukeetafotieputaylatohsightaylatu
keeteemaytieputaylateemaytukeeteemaytohsightukee
tafoteemaytukeetohsighteemaytohsighteemaytohsight
afoteemaytafotohsightukeetieputukeetieputafoteemay
tohsighteemaytukeetafotieputaylatohsightaylatukee
teemaytieputaylateemaytukeeteemaytohsi teemaytoh
sighteemaytukeetafotieputaylatohsightaylatukeetee
maytieputaylateemaytukeeteemaytohsightukeetafotee
maytukeetohsighteemaytohsighteemaytohsightafotee

maytafotohsightukeetieputukeetieputafoteemaytoh
sighteemaytukeetafotieputaylatohsightaylatukeetee
maytieputaylateemaytukeeteemaytohsi teemaytoh
sighteemaytukeetafotieputaylatohsightaylatukeetee
maytieputaylateemaytukeeteemaytohsightukeetafotee
maytukeetohsighteemaytohsighteemaytohsightafotee
maytafotohsightukeetieputukeetieputafoteemaytoh
sighteemaytukeetafotieputaylatohsightaylatukeetee
maytieputaylateemaytukeeteemaytohsi teemaytoh
sighteemaytukeetafotieputaylatohsightaylatukeetee
maytieputaylateemaytukeeteemaytohsightukeetafotee
maytukeetohsi.

Child Development, May/June 2010, Volume 81, Number 3, Pages 737–756

Do Effects of Early Child Care Extend to Age 15 Years? Results From the NICHD Study of Early Child Care and Youth Development

Deborah Lowe Vandell
University of California, Irvine

Margaret Burchinal
University of California, Irvine

Nathan Vandergrift
Duke University

Jay Belsky
Birkbeck University of London

Laurence Steinberg
Temple University

NICHD Early Child Care Research Network

Relations between nonrelative child care (birth to 4½ years) and functioning at age 15 were examined (N = 1,364). Both quality and quantity of child care were linked to adolescent functioning. Effects were similar in size as those observed at younger ages. Higher quality care predicted higher cognitive–academic achievement at age 15, with escalating positive effects at higher levels of quality. The association between quality and achievement was mediated, in part, by earlier child-care effects on achievement. High-quality early child care also predicted youth reports of less externalizing behavior. More hours of nonrelative care predicted greater risk taking and impulsivity at age 15, relations that were partially mediated by earlier child-care effects on externalizing behaviors.

The transition from childhood to adolescence involves substantial changes in multiple features of children's lives, which raises fundamental questions about the importance of early experience as an influence on adolescent development. Adolescence is defined by physical and cognitive changes (Kuhn, 2009; Susman & Dorn, 2009) as well as transformations in parent–child and peer

---

This study is directed by a Steering Committee and supported by NICHD through a cooperative agreement (5 U10 HD027040), which calls for scientific collaboration between the grantees and the NICHD staff. The content is solely the responsibility of the named authors and does not necessarily represent the official views of the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, the National Institutes of Health, or individual members of the Network. Current members of the Steering Committee of the NICHD Early Child Care Research Network, listed in alphabetical order, are: Jay Belsky (Birkbeck University of London), Cathryn Booth-LaForce (University of Washington), Robert H. Bradley (Arizona State University), Celia A. Brownell (University of Pittsburgh), Margaret Burchinal (University of California, Irvine), Susan B. Campbell (University of Pittsburgh), Elizabeth Cauffman (University of California, Irvine), Alison Clarke-Stewart (University of California, Irvine), Martha Cox (University of North Carolina, Chapel Hill), Robert Crosnoe (University of Texas, Austin), Sarah L. Friedman (Institute for Public Research, CNA), James A. Griffin (NICHD), Bonnie Halpern-Felsher (University of California, San Francisco), Willard Hartup (University of Minnesota), Kathryn Hirsh-Pasek (Temple University), Daniel Keating (University of Michigan, Ann Arbor), Bonnie Knoke (RTI International), Tama Leventhal (Tufts University), Kathleen McCartney (Harvard University), Vonnie C. McLoyd (University of North Carolina, Chapel Hill), Fred Morrison (University of Michigan, Ann Arbor), Philip Nader (University of California, San Diego), Marion O'Brien (University of North Carolina, Greensboro), Margaret Tresch Owen (University of Texas, Dallas), Ross Parke (University of California, Riverside), Robert Pianta (University of Virginia), Kim M. Pierce (University of California, Irvine), A. Vijaya Rao (RTI International), Glenn I. Roisman (University of Illinois at Urbana-Champaign), Susan Spieker (University of Washington), Laurence Steinberg (Temple University), Elizabeth Susman (Pennsylvania State University), Deborah Lowe Vandell (University of California, Irvine), and Marsha Weinraub (Temple University).

The authors express their appreciation to the study coordinators at each site who oversaw the data collection, to the research assistants who collected the data, and especially to the youth and their families who cooperated so willingly with repeated requests for information.

Correspondence concerning this article should be addressed to Deborah Lowe Vandell, Department of Education, University of California, 3200 Education Building, Irvine, CA 92697. Electronic mail may be sent to dvandell@uci.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8103-0007

relationships (Collins & Steinberg, 2006) and schooling (Eccles & Roeser, 2009). With these myriad changes, there is reason to wonder whether effects of early child-care experiences persist into adolescence. This is the central issue addressed in this report. Specifically, we ask if nonrelative child care during the first 4½ years of life predicts academic achievement and behavioral adjustment at age 15. Then, we consider developmental processes that may mediate these associations. Finally, we ask if links between early child care and adolescent outcomes are moderated by child gender or familial risk.

The work to be reported is based on a large, nonexperimental field study—the National Institute of Child Health and Human Development Study of Early Child Care and Youth Development (NICHD SECCYD)—that affords estimation of statistical rather than causal effects. When causal language (e.g., effect, influence) is employed in this report, it is for heuristic purposes.

Child Care and Child Development

Two different perspectives have guided much of the research examining the effects of early child care. For at least 50 years, nursery schools and preschools have been viewed by parents and educators as a means to promote social and academic skills prior to entry to formal schooling (Lamb & Ahnert, 2006). In contrast, others, influenced in part by attachment theory, have expressed concerns that extensive nonmaternal care, especially beginning very early in life, could disrupt attachment bonds and result in problem behaviors (Belsky, 1986, 1988; Egeland & Hiester, 1995).

Research findings provide support for both views. Experimental studies of high-quality early intervention programs have demonstrated that these programs can enhance social, cognitive, and academic development of economically disadvantaged children (Campbell, Pungello, Miller-Johnson, Burchinal, & Ramey, 2001; Love et al., 2005; Reynolds, 2000; Schweinhart, Weikart, & Larner, 1986). Correlational studies of economically and ethnically diverse samples also have fairly consistently found higher quality child care to be associated with better cognitive and academic outcomes (Broberg, Wessels, Lamb, & Hwang, 1997; Burchinal et al., 2000; Cote et al., 2007; Gormley, Gayer, Phillips, & Dawson, 2005; Mashburn et al., 2008; Peisner-Feinberg & Burchinal, 1997). Evidence of social benefits of child care has been more mixed. Although some have found benefits of high-quality care for social development (Cote et al., 2007; Howes, Phillips, & Whitebook, 1992; Vandell, Henderson, & Wilson, 1988), other researchers have identified potentially adverse consequences of long hours of care, especially, though not exclusively, if initiated early in life (Bates et al., 1994; Belsky, 2001; Cote, Borge, Geoffroy, Rutter, & Tremblay, 2008; Haskins, 1985; Loeb, Bridges, Bassok, Fuller, & Rumberger, 2007; Nomaguchi, 2006; Vandell & Corasaniti, 1990). Time in center-type settings has been related to negative social behavioral outcomes but also positive academic outcomes (Huston et al., 2001; Loeb et al., 2007; Magnuson, Ruhm, & Waldfogel, 2007). Efforts to understand and integrate these disparate findings has led to a conceptualization of child care that differentiates quality of care, quantity of care, and types of care as potentially distinct influences on children's development.

The NICHD SECCYD was launched in the early 1990s to examine the effects of these three distinctive aspects of early child care. In previous reports of child functioning prior to school entry (NICHD Early Child Care Research Network [ECCRN], 2002), in the primary grades (NICHD ECCRN, 2005c), and later in elementary school (Belsky et al., 2007), child-care quality, quantity, and type were differentially linked to children's development. In general, higher quality of child care was related to higher cognitive–academic performance, whereas more hours of child care (especially by nonrelatives) was related to more problem behavior. More experience in center-type care was related to better cognitive skills but also more problem behavior. Although this general pattern of findings was detected at three time periods (preschool, early elementary, and later elementary), effects were small by traditional standards, and over time some previously detected child-care effects disappeared when evaluated at later ages, raising the core issue of the current inquiry, namely, whether effects of early child care are evident in adolescence.

Some have contended that child-care effects would fade away over time (Blau, 1999; Colwell, Pettit, Meece, Bates, & Dodge, 2001; Deater-Deckard, Pinkerton, & Scarr, 1996), especially by the time young people are in high school because subsequent life experiences likely override child-care experiences that occurred a decade or more earlier. At the same time, however, in view of the fact that development in adolescence builds on prior periods and that some effects of early experience may not manifest themselves until adolescence (so-called sleeper effects), some effects of early child care might remain even among teenagers.

To date, only a few early intervention studies have tracked effects of high-quality early child care and education into the high school years and beyond. Low-income children exposed to high-quality program care show better academic outcomes through high school and higher rates of employment and less criminal activity as young adults (Campbell et al., 2001; Lazar & Darlington, 1982; Schweinhart et al., 1986). Several large longitudinal studies that included middle- as well as low-income participants have documented quality effects in elementary school (Melhuish et al., 2008; Peisner-Feinberg et al., 2001), but none has considered child-care effects in adolescence.

This study extends previous research by examining the links between routine early child-care experience (i.e., not high-quality early intervention programs) and adolescent functioning at age 15 in a large and economically diverse sample. Three related issues are considered. The first concerns the specificity of child-care effects. Previous analyses of SECCYD data indicated that child-care quality positively predicts cognitive and academic functioning, more hours of care predicts more problem behavior, and more experience in center care predicts both better academic outcomes and more problem behavior at various times during the preschool and middle childhood years (Belsky et al., 2007; NICHD ECCRN, 1998, 2003, 2005c). Here, we examine whether such domain specificity is maintained at age 15.

A second issue involves possible pathways through which child care could affect adolescent development. The most simple and straightforward proposition is that differences in child functioning at entry to school that are linked to early care are carried forward to middle adolescence. We address this possibility by evaluating the extent to which observed effects of child care on adolescent functioning at age 15 are mediated by prior cognitive and social functioning in early and middle childhood.

A third issue that merits attention is whether associations between child care and adolescent functioning are moderated by child gender or familial risk. Hypotheses regarding differential gender effects have been in the literature for years (Belsky, 1988; Love et al., 2003; Maccoby & Lewis, 2003), with some evidence of these effects reported for child care and developmental outcomes in the preschool and early elementary years (Caughy, DiPietro, & Strobino, 1994; Crockenberg, 2003; Peisner-Feinberg et al., 2001). Additionally, some early studies of maternal employment highlighted negative effects for boys and positive effects for girls (Gottfried & Gottfried, 1988; Hoffman & Youngblade, 1999). A recent reanalysis of the Abecedarian, Perry Preschool, and the Early Training Projects found long-term educational benefits for girls, but not boys (Anderson, 2008), although a cost–benefit analysis of the Perry Preschool participants at the age of 40 indicated greater benefits for men than women, primarily because of economic savings associated with reductions in the men's incarcerations (Belfield, Nores, Barnett, & Schweinhart, 2006).

There also is some evidence that the effects of child care may vary as a function of familial risk. Quality of early child care has emerged as a protective factor of familial social risk (measured by maternal education, family income, household size, and maternal depression) in terms of academic achievement in elementary school (Burchinal, Vandergrift, & Pianta, 2009). A report based on a large national survey found that children of low-income families benefited more in terms of cognitive–academic performance when they experienced more hours of child care (labeled a compensatory effect), whereas children of middle-income households functioned more poorly when they experienced more hours of child care (labeled a lost-resources effect; Desai, Chase-Lansdale, & Michael, 1989). In other research, more time in center care predicted larger academic gains among low-income than middle-income children (Gormley et al., 2005; Magnuson et al., 2007). For the most part, however, previous analyses of the SECCYD have failed to detect familial risk- or gender-moderated child-care effects (NICHD ECCRN, 2002, 2005c). We examine whether such effects emerge in adolescence.

In all nonexperimental studies of child care, selection bias is a concern because family and child characteristics are related to quality, type, and hours of care as well as to children's functioning (Committee on Family and Work Policies, 2003). To reduce the likelihood of selection bias, extensive covariates measured in early childhood are used as control variables in all analyses, following the practice adopted in our previous reports. Primary analyses also include extensive covariates measured in middle childhood and adolescence, in line with our prior work and consistent with the standard adopted by developmental psychologists who have argued that controlling for subsequent family and school experiences provides a more conservative test of effects associated with early experience (Bradley, Caldwell, & Rock, 1988;

Kovan, Chung, & Sroufe, 2009; Lamb, Thompson, & Gardner, 1984). Failing to include such controls risks attributing effects to early experience that could just as well (and perhaps more parsimoniously) be a function of later experience. Nevertheless, we also test relations without the middle childhood and adolescent covariates because family factors in middle childhood and adolescence may have been influenced by child care. If early child care affected these covariates, then their inclusion might result in over- or underestimates of child-care effects.

In sum, this report provides new insights into potential long-term child-care effects by tracking a large sample of American children to 15 years of age to determine whether variations in early child-care quality, quantity, and type are related to cognitive development, academic achievement, and socioemotional functioning in adolescence.

Method

Participants

Hospital visits were conducted with mothers shortly after the birth of a child in 1991 in 10 locations in the United States. During selected 24-hr intervals, all women giving birth ($N = 8{,}986$) were screened for eligibility. Of those families, 3,142 were excluded owing to a priori criteria such as failure to speak English and plans to move within the next 3 years. At a follow-up telephone interview at 2 weeks, 1,353 could not be contacted or refused to participate. Families were randomly selected among the remaining pool of eligible participants. A total of 1,364 families were recruited, completed a home interview at 1 month, and became the study participants. Overall, this constituted a 52% response rate from the original approach to families in the hospital to successful recruitment in the study.

In terms of demographic characteristics, 26% of the mothers in the recruited sample had no more than a high school education at recruitment, 21% had incomes no greater than 200% of the poverty level, and 22% were minority (i.e., not non-Hispanic Euro-American). Details of the sampling plan can be found in NICHD ECCRN (2005b). It should be noted that the primary analyses to be presented estimate effects based on all 1,364 children originally recruited into the sample.

At age 15, measures of adolescent outcomes were obtained for 958 youth (70% of the original recruitment sample). Comparisons of the age 15 sample and the other 406 youth in the birth cohort sample revealed four significant differences on the 30 early and middle childhood variables listed in Tables 1 and 2. Nonparticipants were more likely to be boys (56% vs. 50%) and have lower scores at 4½ years on a test of math skills (97.8 vs. 102.5), and their mothers were less educated (13.4 years vs. 14.3 years) and provided lower quality parenting (−.25 standardized parenting score vs. −.02 standardized parenting score).

Measures

Children were studied from birth to age 15. Assessments occurred when the children were 1, 6, 15, 24, 36, and 54 months old; when they were in kindergarten and Grades 1, 2, 3, 4, 5, and 6; and at age 15. The following sections describe the specific measures used in the present analyses and the time points of administration. Additional details about all data collection procedures, psychometric properties of the instruments, and descriptions of how composites were derived and constructed can be found in the study's Manuals of Operation and Instrument Documentation (http://secc.rti.org).

Measurements are described in terms of their roles in the analyses. Measures reflecting the child's experiences in early child care are presented first, followed by the adolescent cognitive and social outcome measures. Variables used to control for family factors and schooling are then described. Finally, we describe measures of child functioning in early and middle childhood that are used in the mediation pathway analyses.

Child-Care Characteristics

Three aspects of child care were measured from birth through 4½ years: type of care, quantity of care, and quality of care (see Table 1).

Child-care type. During telephone and personal interviews conducted at 3-month intervals (or epochs) through 36 months and 4-month intervals (or epochs) to 4½ years of age, mothers reported types and hours of regularly used nonmaternal care. During each interview, mothers reported all of the care arrangements used since the previous interview. Arrangements were classified as center, child-care home (any home-based care outside the child's own home), in-home care (any caregiver in the child's own home), father care, and grandparent care. The proportion of epochs in which the child received care in a center for at least 10 hr/week was computed and used as variable to represent *type of care*.

Table 1
Descriptive Statistics for Child-Care Variables and Covariates

	Early childhood	Middle childhood	Adolescence
Child care			
Hours			
N	1,214		
M	16.48		
(*SD*)	(14.16)		
Quality			
N	1,005		
M	2.90		
(*SD*)	(0.45)		
% center			
N	1,214		
M	0.21		
(*SD*)	(0.26)		
Covariate			
Male			
N	1,364		
%	52		
Race (*N*)	1,364		
Black (%)	13		
Hispanic (%)	6		
White (%)	76		
Other (%)	5		
Maternal education			
N	1,363		
M	14.23		
(*SD*)	(2.51)		
Maternal PPVT–R			
N	1,167		
M	99.01		
(*SD*)	(28.35)		
Maternal adjustment			
N	1,272		
M	59.00		
(*SD*)	(13.95)		
Income-to-needs ratio			
N	1,302	1,140	924
M	3.60	4.19	5.26
(*SD*)	(2.85)	(3.37)	(5.79)
% of epochs 2-parents			
N	1,305	1,154	979
M	84	81	77
(*SD*)	(32)	(34)	(42)
Maternal depression			
N	1,304	1,123	973
M	9.36	8.74	10.48
(*SD*)	(6.76)	(7.32)	(9.83)
Parenting composite			
N	1,306	1,306	1,306
M	−0.03	−0.03	−0.03
(*SD*)	(0.73)	(0.73)	(0.73)
Classroom quality composite			
N		1,100	
M		3.27	
(*SD*)		(5.79)	

Note. PPVT–R = Peabody Picture Vocabulary Test–Revised.

Table 2
Descriptive Statistics for Child Outcomes at Age 15 and Earlier Ages

		Measurement period			
Child outcome	Age 15	4½ years	Grade 1	Grade 3	Grade 5
Cognitive–academic composite					
N	892	1,060	1,026	1,016	993
M	106.1	100.6	109.4	110.9	107.2
(*SD*)	(13.1)	(12.5)	(13.1)	(13.2)	(13.2)
WJ Vocabulary					
N	889	1,060	1,020	1,014	992
M	99.9	100.2	105.5	105.5	103.9
(*SD*)	(14.8)	(15.3)	(15.6)	(14.8)	(14.8)
WJ Reading[a]					
N	887	1,056	1,025	1,011	993
M	107.7	98.9	120.0	111.1	107.9
(*SD*)	(15.7)	(13.5)	(15.8)	(14.0)	(13.9)
WJ Math[b]					
N	887	1,053	1,023	1,012	993
M	102.9	102.9	110.8	116.3	110.7
(*SD*)	(14.2)	(15.6)	(17.1)	(17.3)	(17.4)
WJ Analogies					
N	891				
M	113.7				
(*SD*)	(16.0)				
Externalizing[c]					
N	956	705	1,007	982	927
M	49.3	50.2	50.7	51.5	51.0
(*SD*)	(9.9)	(9.6)	(8.7)	(9.4)	(9.2)
Risk taking					
N	954				
M	37.63				
(*SD*)	(19.5)				
Impulsivity					
N	957				
M	−35.1				
(*SD*)	(9.0)				

[a] Reading measures were Woodcock–Johnson (WJ) Passage Comprehension at age 15, Broad Reading in Grades 3 and 5, and Letter-Word Identification at 4½ years and Grade 1.

[b] Math measures were WJ Applied Problems at 4½ years, Grade 1, and age 15, and Broad Math in Grades 3 and 5.

[c] Externalizing measure was Youth Self-Report at 15 years and Teacher Report at all other ages.

Child-care hours. The hours per week in all types of nonmaternal care excluding fathers and grandparents were tallied for each epoch, and then the mean of nonrelative care hours across epochs was computed.

Child-care quality. Observational assessments were conducted in the primary child-care arrangement at ages 6, 15, 24, 36, and 54 months. Quality was assessed during two half-day visits scheduled within a 2-week interval at 6–36 months and one half-day visit at 54 months. Observers completed four 44-min cycles of the Observational Record of the Caregiving Environment (ORCE) per child age through 36 months and two 44-min ORCE cycles at 54 months. Detailed descriptions of the ORCE assessments can be found in NICHD ECCRN (2002), including coding definitions, training procedures, and interobserver agreement. Reliability exceeded .90 at 6 months, .86 at 15 months, .81 at 24 months, .80 at 36 months, and .90 at 54 months.

A mean quality of nonrelative care score was computed for each child.

Adolescent Functioning at Age 15

Table 2 provides the descriptive statistics for the measures of adolescent functioning at age 15.

Cognitive–academic achievement. The Woodcock–Johnson Psycho-Educational Battery–Revised (WJ–R; Woodcock & Johnson, 1989) is a wide-range, comprehensive set of individually administered tests that consists of two major parts: the Tests of Cognitive Ability and the Tests of Achievement. At age 15, cognitive ability was assessed with two subscales, Picture Vocabulary and Verbal Analogies. Achievement was assessed using the Passage Comprehension and Applied Problems subscales. In this report, standard scores, which are based on a mean of 100 and a standard deviation of 15, and the equivalent percentile rank were used.

Risk taking. Adolescents reported *risk-taking* behaviors using an audio computer-assisted self-interview. Thirty-six risk-taking items were drawn from instruments used in prior studies of adolescents (Halpern-Felsher, Biehl, Kropp, & Rubinstein, 2004). Adolescents reported the extent to which, over the past year, they used alcohol, tobacco, or other drugs; behaved in ways that threatened their own safety (e.g., rode in a vehicle without the use of seatbelts); used or threatened to use a weapon; stole something; or harmed property. Responses were made on a 3-point scale (0 = *never,* 1 = *once or twice,* 2 = *more than twice*). Ratings were summed across component items and then subjected to square-root transformation to reduce skew and kurtosis (Cronbach alpha = .89).

Impulsivity. Adolescents completed an eight-item questionnaire to assess reactions to external constraints, taken from the Weinberger Adjustment Inventory (Weinberger & Schwartz, 1990). The measure asks participants to rate (1 = *false* to 5 = *true*) how closely their behavior matched a series of statements. Sample items include: "I'm the kind of person who will try anything once, even if it's not that safe," "I should try harder to control myself when I'm having fun," and "I do things without giving them enough thought." Seven items were used to create an impulsivity composite score (Cronbach alpha = .82).

Externalizing problems. Adolescents self-reported externalizing behaviors using the Youth Self-Report (YSR; Achenbach & Rescorla, 2001). The scale consists of 119 items that reflect a broad range of behavioral and emotional problems as well as 16 socially desirable items. For each item, the adolescent is asked to rate how well that item describes him or her currently or within the last 6 months: on a 3-point scale (0 = *not true,* 1 = *somewhat or sometimes true,* 2 = *very true or often true*). The format of the YSR is similar to that of the Child Behavior Checklist completed by parents (CBCL; Achenbach, 1991a) and the Teacher Report Form (TRF) completed by teachers (Achenbach, 1991b). The YSR, CBCL, and TRF have 89 problem items in common, but the YSR includes additional items that are specifically designed for adolescents. *Externalizing behaviors* are assessed by 30 items (Cronbach alpha = .86).

Maternal, Child, Family, and School Controls

Measures of maternal, child, and family characteristics during early childhood, middle childhood, and adolescence, and school quality were used as controls for possible selection bias. See Table 1 for the descriptive statistics.

The early childhood covariates are *maternal education* (in years); child *gender;* child *race and ethnicity;* the proportion of epochs through 4½ years in which the mother reported a *husband* or *partner* was present; family income through 4½ years calculated as the mean *income-to-needs ratio; maternal Peabody Picture Vocabulary Test–Revised* (PPVT–R; Dunn & Dunn, 1981) obtained when the study child was 3 years of age; *maternal psychological adjustment* measured when the study child was 6 months of age using three subscales (Neuroticism, Extraversion, and Agreeableness) of the NEO Personality Inventory (Costa & McCrae, 1985, 1989); the mean of *maternal depressive symptoms* assessed by the Center for Epidemiological Studies Depression Scale (CES–D; Radloff, 1977) reported by the mother at 6, 15, 24, 36, and 54 months; and an *early parenting quality* composite score created by averaging standardized ratings of observed maternal sensitivity and observed home environmental quality measured at 6, 15, 24, 36, and 54 months. These control variables are described in detail in NICHD ECCRN (2002) and the instrument documentation available at the project Web site (http://secc.rti.org).

The middle childhood covariates were measured when the study children were in Grades 1, 3, and 5. These family covariates paralleled those obtained in early childhood: the proportion of the middle childhood in which a *husband or partner* was present in the household, mean *income-to-needs ratio*, mean *maternal depressive symptoms*, and *mean observed parenting quality* (the average of standardized ratings of maternal sensitivity in a semistructured task and a home observation). In addition, the quality of school experiences in middle childhood was rated during two 44-min observations in Grade 1 and eight 44-min observations in Grades 3 and 5, and a *mean classroom quality score* was computed. For more details about the classroom observations and the classroom quality composites, see NICHD ECCRN (2004, 2005a).

The adolescent family covariates collected at age 15 corresponded to those used in early and middle childhood: *presence of a husband or partner* in the household, *income-to-needs ratio*, *maternal depressive symptoms*, and the *observed parenting quality composite*.

The age 15 parenting quality composite was based on ratings of maternal sensitivity made from a video-recorded 8-min discussion of areas of disagreement between the adolescent and mother (e.g., chores, homework, and money), selected by the adolescent (Allen et al., 2000), and a home observation combined with a semistructured interview (HOME; Bradley et al., 2000). The discussion task was coded using 7-point rating scales, with higher scores indicating higher levels of sensitivity based on adaptations of the more microanalytic coding systems of Allen et al. (2000, 2001) and coding systems used at earlier ages in the SECCYD (NICHD ECCRN, 2008). Maternal sensitivity was the sum of the 7-point ratings of supportive presence, respect for autonomy, and hostility (reversed). Cronbach alphas for the sensitivity composite scores ranged from .80 to .85 and interrater reliabilities determined from intraclass correlations based on a second coding of 19.5% (196/1004) to 27% (271/987) of the videotapes at different ages ranged from .84 to .91.

The adolescent version of the HOME scale consists of 44 items that assess five domains (Physical Elements, Learning Materials, Variety of Experiences, Acceptance and Responsiveness, and Regulatory Activities), which are combined to create a total score. The maternal sensitivity and HOME scores were standardized and averaged to create the parenting quality composite.

Child Functioning in Early and Middle Childhood

Measures of children's functioning in early and middle childhood included cognitive–academic achievement and behavior problems. Descriptive statistics for these measures can be found in Table 2.

For *cognitive–academic achievement*, children were administered subtests from the WJ–R: *Letter-Word Identification* (4½ years and first grade) and *Broad Reading* (third and fifth grades), which adds assessment of passage comprehension to the assessment of identification of words; *Applied Problems*, which measures skill in analyzing and solving practical problems in mathematics; and *Picture Vocabulary*, which measures children's ability to name objects depicted in a series of pictures.

The TRF (Achenbach, 1991b) was used to evaluate *Externalizing Problems* (e.g., ''hits others,'' ''disobedient at school,'' and ''argues a lot'') at 4½ years and in first, third, and fifth grades.

Results

Data analyses focus on whether early child-care quality, quantity, and type are associated with adolescent outcomes at 15 years of age, and if so, whether child functioning at entry to school mediates these subsequent associations. We also examine whether gender or family risk moderate associations between early child care and age 15 outcomes. Descriptive analyses are conducted and structural equation models (SEMs) test direct pathways, mediated pathways, and moderated relations between early child-care experiences and 15-year outcomes. The SEMs account for missing data through full-information maximum likelihood (FIML) based on the data for all 1,364 children originally recruited for the study. Alphas were set at .05, and thus all results reported are significant at $p < .05$ or better.

Preliminary Analyses

Descriptive statistics are presented in Tables 1–3. Table 1 describes the sample, presenting percentages, means, and standard deviations for the child-care variables and the selected covariates in the early childhood, middle childhood, and adolescent periods. Figure 1 shows the sample distributions for child-care quality, quantity, and type. Seventeen percent of the children experienced high-quality child care (ORCE scores of 3.30 or higher, on average), and 24% experienced moderately high-quality nonrelative care (ORCE scores of 3.0–3.29 on average). Less than one fourth of the children (21%) were in nonrelative care for more than 30 hr/week. Sixty-four percent of the children participated in center-type care for at least one epoch, with 24% experiencing more than 1 year of center care by 4½ years of age.

Table 2 provides descriptive statistics for the adolescent outcomes and for children's prior academic and behavioral functioning. The first column lists the means and standard deviations for cognitive–academic and behavioral outcomes at age 15. The next four columns show earlier assessments of cognitive–academic functioning and teacher-reported externalizing problems that are examined as potential pathways by which child-care experiences might be linked to age 15 outcomes. Table 3 presents correlations between child-care variables and all child outcomes. Because the analysis created a latent cognitive–academic achievement variable, a manifest composite was created for these correlations.

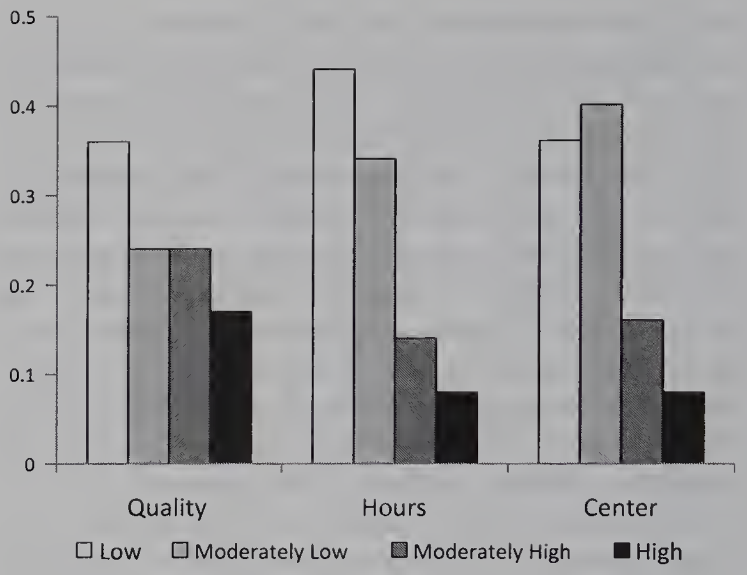


Figure 1. Distribution of quality, hours, and type of nonrelative child care.
Note. Quality categories: low: Observational Record of the Caregiving Environment (ORCE) ≤ 2.75; moderately low: ORCE = 2.75 to < 3.0; moderately high: ORCE = 3.00 to < 3.30; high: ORCE = 3.30–4.00. Hours categories: low: < 10 hr; moderately low: 10 to < 30 hr; moderately high: 30–40; high: > 40. Center categories: low: 0; moderately low: > 0 but less than 33%; moderately high: 33% to < 67%; high: 67%–100%.

Table 3
Correlations Among Child-Care Variables and Child Outcomes

Variable	Hours	Child-care quality	Center
Child care			
Hours	1.00		
Quality	−0.20	1.00	
Center	0.51	−0.18	1.00
15-year outcomes			
Achievement age 15	0.04	0.23	0.05
Externalizing age 15	0.02	−0.11	−0.04
Impulsivity age 15	0.08	−0.10	0.02
Risk-taking age 15	0.04	−0.12	0.02
Hypothesized pathways			
Achievement 4½ years	0.10	0.22	0.08
Achievement G1	0.11	0.16	0.11
Achievement G3	0.09	0.15	0.07
Achievement G5	0.08	0.18	0.05
Externalizing 4½ years	0.20	−0.17	0.16
Externalizing G1	0.17	−0.11	0.11
Externalizing G3	0.10	−0.15	0.10
Externalizing G5	0.05	−0.11	0.04

Note. G1 = Grade 1; G3 = Grade 3; G5 = Grade 5.

Substantive Analyses

SEMs with FIML are used to test three sets of issues. The first examines direct associations between early child-care experiences and adolescent outcomes at age 15. The second adds potential mediators and the third tests potential moderators. The use of FIML allows the inclusion of the entire sample ($N = 1{,}364$) in these analyses. FIML has been shown to be as effective as multiple imputation in addressing problems associated with missing data (Schafer & Graham, 2002). When statistically significant associations are found, effect sizes are computed as the anticipated change in the outcome in standard deviation units when the predictor changes by a standard deviation or $d = B \times SD_{\text{predictor}}/SD_{\text{outcome}}$ (see NICHD ECCRN & Duncan, 2003).

Direct Associations Between Early Child Care and Age 15 Outcomes

The first set of analyses fit an SEM in which age 15 outcomes were predicted from early child-care hours, type, and quality, adjusting for family experiences in early and middle childhood and adolescence and school experiences in middle childhood (see Figure 2). The particular child-care variables of interest were child-care hours, proportion of time in center care, and child-care quality, computed from all care settings except those that involved fathers or grandparents. Preliminary analyses tested whether associations between the child-care variables and age 15 outcomes were linear or quadratic, indicating that only quality showed the quadratic association and that having the other quadratic terms did not increase fit, likelihood-ratio test $\chi^2(6) = 3.62$, $p = .27$. Accordingly, the quality-squared term was added to the model.

Academic achievement at age 15 was treated as a latent construct, with indicators of WJ–R Applied Problems, Picture Vocabulary, Passage Comprehension, and Verbal Analogies scores. Preliminary attempts to form a latent construct for problem behaviors suggested a poor fit, so adolescent self-reports of externalizing, risk taking, and impulsivity were analyzed as separate manifest variables. The model also included the following covariates: research site, child gender, child ethnicity, maternal education, maternal PPVT, maternal psychological adjustment, elementary school classroom quality, and repeated assessments from early childhood, middle childhood, and adolescence of family income, proportion time mother had a husband or partner, maternal depressive symptoms, and parenting quality. Including family covariates from the three developmental periods (early childhood, middle childhood, and adolescence) reduced the likelihood that observed associations between child outcomes and early child-care experience were because of family selection factors.

The SEM model fit the data well, $\chi^2 = 237.52(113)$; $\chi^2/df = 2.10$; comparative fit index (CFI) = .97; root mean square error of approximation (RMSEA) = .03. Generally, an RMSEA < .10 and a CFI ≥ .90 indicate good fit (Bollen, 1989). Table 4 shows the estimated coefficients and effect sizes for the paths from child-care quality variables to age 15 outcomes, and Figure 2 displays the statistically significant standardized coefficients. Results indicated that child-care quality showed significant linear ($B = 2.62$, $d = .09$) and quadratic ($B = 3.35$, $d = .07$) associations with children's

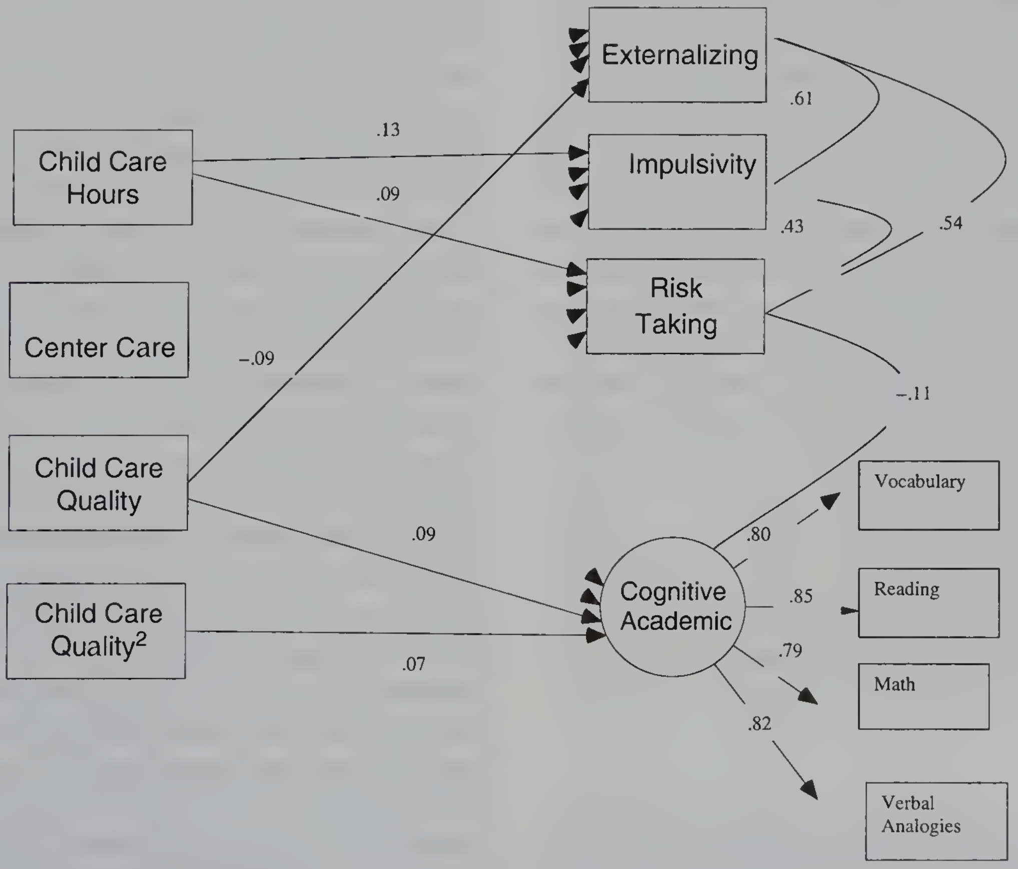


Figure 2. Standardized path coefficients in structural equation models relating child care experiences to 15-year outcomes.
Note. Only statistically significant paths are shown. Covariates include site, gender, ethnicity, maternal education, maternal Peabody Picture Vocabulary Test–Revised, maternal adjustment, elementary school classroom quality, and repeated assessments from early childhood, middle childhood, and adolescence of family income, proportion of time mother had a husband or partner, maternal depressive symptoms, and parenting quality.

Table 4
Path Coefficients From Structural Equation Model Relating Child-Care Experiences to 15-Year Outcomes

	Cognitive–academic achievement			Externalizing			Risk taking			Impulsivity		
	B	*(SE)*	*d*	*B*	*(SE)*	*d*	*B*	*(SE)*	*d*	*B*	*(SE)*	*d*
Hours	0.02	(0.03)	.03	0.03	(0.03)	.04	**0.008***	**(0.003)**	**.09**	**.08*****	**(0.03)**	**.13**
Center care	1.11	(1.74)	.02	−0.95	(1.51)	.02	−0.15	(0.18)	−.03	−1.39	(1.25)	−.04
Quality	**2.62*****	**(0.99)**	**.09**	**−1.89***	**(0.86)**	**−.09**	−0.10	(0.10)	−.04	−0.97	(0.78)	−.05
Quality-squared	**3.35****	**(1.50)**	**.07**	−0.78	(1.30)	−.02	−0.25	(0.15)	−.06	−0.92	(1.17)	−.03

Note. Model also includes as covariates site, gender, maternal education, maternal Peabody Picture Vocabulary Test–Revised, maternal adjustment, early childhood risk index, ethnicity, elementary school classroom quality and repeated assessments of family income, proportion time mother had husband/partner, maternal depressive symptoms, and parenting. A square-root transformation was applied to risk taking owing to its skewed distribution. *d* is analogous to the effect size measure Cohen's *d*, computed as $B \times SD_{\text{predictor}}/SD_{\text{outcomes}}$.
$^{*}p < .05.$ $^{**}p < .01.$ $^{***}p < .001.$

cognitive–academic achievement at age 15. The linear association suggested that children who experienced higher quality care had significantly higher levels of cognitive–academic achievement at age 15 whereas the quadratic association indicated that associations were stronger at moderately high levels of quality than at low or very low levels.

Figure 3 shows the estimated nonlinear association between child-care quality and achievement. Plotted are predicted values from −2.0 to +2.0 *SD* around the quality mean of 2.90. As shown, higher quality is associated with higher academic outcomes when ORCE scores are 2.75 or higher. In contrast, quality is nonsignificantly negatively related to academic outcomes when mean quality is in the very low range, averaging less than 2.5 on the ORCE over the first 4½ years. The turning point or minimum point of the quadratic function describing the association between quality and academic outcomes is computed by taking the derivative of the quadratic function. This value is −.39 on the mean-centered scale and 2.51 in the original ORCE scale. Thus, the turning point in this function is almost 1 *SD* (0.45) below the sample mean (2.90). The effect sizes for quality vary around this turning point. To illustrate, effect sizes were computed in terms of standard deviation units of child-care quality: *d* = .193 at +2 *SD* (ORCE = 3.8); *d* = .168 at +1.5 *SD* (ORCE = 3.58); *d* = .142 at +1 *SD* (ORCE = 3.35), *d* = .116 at +0.5 *SD* (ORCE = 3.13), *d* = .090 at the mean (ORCE = 2.90), *d* = .064 at −0.5 *SD* (ORCE = 2.67), *d* = .038 at −1 *SD* (ORCE = 2.45), and *d* = −.039 at −2 *SD* (ORCE = 2.00).

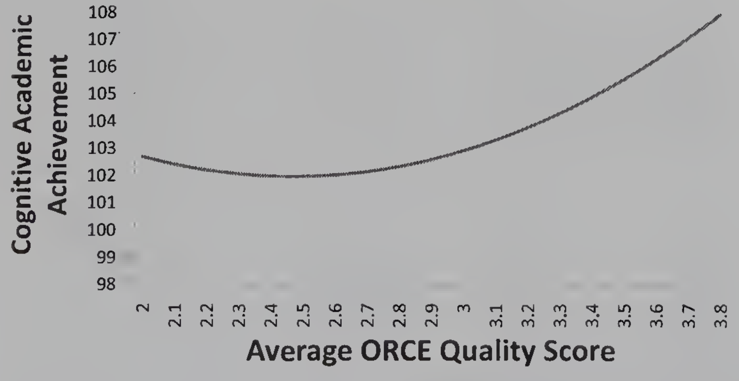


Figure 3. Predicting cognitive–academic achievement from child-care quality using quadratic regression.
Note. ORCE = Observational Record of the Caregiving Environment.

Table 4 and Figure 2 also show significant direct pathways between early child-care experience and problem behaviors. Adolescents who experienced more hours of nonrelative child care across their first 4½ years reported significantly more risk taking (*B* = 0.008, *d* = .09) and greater impulsivity (*B* = 0.08, *d* = .13) at age 15. In addition, children who experienced higher quality care had significantly lower externalizing scores (*B* = −1.89, *d* = .09).

Exposure to center care was not related to academic achievement or problem behaviors at age 15. Several analyses were conducted to examine the robustness of these findings. First, because child hours and proportion time in center care were correlated (*r* = .51), the SEM analysis was recalculated twice, first including quality and type and all covariates (but omitting hours of care) and, second, including quality and hours and all covariates (but omitting type of care). None of the paths involving type was significant in the first follow-up model. The findings associated with hours in these follow-up analyses remained the same as those reported in Table 4. Thus, more hours in nonrelative care, not proportion of measurement occasions spent in centers, predicted poorer social adjustment in adolescence.

Second, the basic SEM analysis examining quality, quality-squared, type, and hours was recalculated including the early childhood covariates and excluding the middle childhood and adolescent covariates. If early child care affects the middle childhood and adolescent covariates, then their inclusion might result in under- or overestimates of the child-care effects. For the most part, the same pattern of significant results was obtained when the middle childhood and adolescent covariates were removed from the model. Quality and quality-squared predicted higher academic outcomes ($B = 2.74$, $SE = 1.0$, $p < .05$, and $B = 3.63$, $SE = 1.51$, $p < .01$, respectively) and quality predicted less externalizing ($B = -1.96$, $SE = 0.87$, $p < .05$). Hours of care predicted more impulsivity ($B = 0.07$, $SE = 0.03$, $p < .001$).

Removing the middle childhood and adolescent covariates resulted in slightly different findings pertaining to adolescent reports of risk taking. More hours predicted more risk taking when all of the covariates were in the model ($p < .05$) but was only marginally significant when the middle childhood and adolescent covariates were excluded ($B = 0.007$, $SE = 0.003$, $p < .06$). Higher quality child care predicted less risk taking when the middle childhood and adolescent covariates were excluded ($B = -0.33$, $SE = 0.15$, $p < .05$) but was only marginally significant when these covariates were included ($B = -0.25$, $SE = 0.15$, $p < .10$).

Third, the quadratic association between child-care quality and cognitive–academic skills was examined in additional analyses to determine the robustness of these findings. A spline approach tested the extent that the association between child-care quality and cognitive–academic skills differed at varying levels of quality. The coefficient for quality as a continuous variable was estimated in models that included two, three, and four quality groups. These models did not provide a good fit to the data (CFI < .90) but did suggest that care quality was a nonsignificantly stronger predictor when quality was in the high range. A dummy variable approach estimated adjusted means for the outcome for the four quality groups. All these analyses suggested that children's cognitive–academic skills at 15 years were higher when they experienced higher quality care and that differences among the group means were larger at the higher end of quality than at the lower end. The two-group spline model with a knot at the mean, ORCE caregiving sensitivity = 2.90, indicated that quality was significantly related to achievement in the higher quality range ($B = 4.61$, $SE = 1.84$, $p = .01$, $d = .12$), was not significantly related in the lower quality range ($B = -0.22$, $SE = 1.60$, $p = .89$, $d = -.01$), and the difference in the magnitude of the association was "marginally" different ($B = 4.84$, $SE = 2.91$, $p = .097$, $d = .14$).

Other analyses that looked at other ORCE values between 2.5 and 3.1 to define higher or lower quality groups did not yield substantially different findings, suggesting that our data may not be able to identify a single cut-point for defining thresholds. These follow-up analyses indicate that the quadratic approach provided the most parsimonious description of the nonlinear association between child-care quality and cognitive–academic outcomes.

Because we had not detected the quadratic association between quality and cognitive–academic achievement at younger ages in our analyses that utilized somewhat different sets of covariates and analytic strategies, we reexamined possible links at these younger ages (54 months, Grades 1, 3, and 5) using the SEM approach and covariates that were used in the current analyses. No significant relations between quality-squared and cognitive–academic achievement were detected at the younger ages in these reanalyses.

Finally, the association between higher child-care quality and lower externalizing problems had not been detected in previous analyses at 54 months or elementary school (Belsky et al., 2007; NICHD ECCRN, 2002, 2003, 2005c), but those analyses also used somewhat different covariates and analytic methods. Therefore, follow-up analyses looked at externalizing ratings by the teacher at 4½ years, and in Grades 1, 3, and 5 using the SEM approach and covariates described earlier. The SEM reanalysis detected one significant relation: Quality of nonrelative care was a quadratic predictor of teacher reports of child externalizing behavior at Grade 1 ($B = -0.07$, $p = .02$). No other significant linear or quadratic associations between child-care quality and children's externalizing were detected between 4½ years and Grade 6.

Prior Functioning as Mediators

Next we sought to identify the pathways that might account for the associations between early child care and adolescent outcomes reported above. These analyses tested our hypothesis that observed associations between child-care experiences and adolescent outcomes are mediated by earlier child care effects. As a result of the complexity of the models, separate SEMs examined the pathways model for cognitive–academic achievement and the pathways model for the behavior outcomes. For these analyses, the initial SEM described in Figure 2 was modified to include child functioning at 4½ years, Grades 1, 3, and 5. Paths from the child-care variables to both the 4½- and 15-year outcomes were included as well as paths between repeated measures of the child functioning from 4½ years to Grade 5, allowing us to estimate a direct path from child-care experiences to age 15 outcomes and a mediated path through the repeated assessments of prior functioning in that outcome domain.

The first pathway model tested the extent to which child-care quality was related to age 15 academic achievement through the earlier associations between child-care quality and academic achievement beginning at 4½ years. Figure 4 displays the standardized paths involving child-care experiences and cognitive–academic achievement estimated by this model. The model fit these data adequately, $\chi^2 = 2447.1(645)$; $\chi^2/df = 3.70$; CFI = .88; RMSEA = .045. The indirect path between child-care quality and academic achievement at age 15 through academic achievement prior to and during elementary school was statistically significant for child-care quality ($B = 1.23$, $SE = 0.57$, $p < .05$, $d = .04$). In contrast, the corresponding direct pathways were statistically nonsignificant. This suggests that the observed association between child-care quality and age 15 cognitive–academic achievement can be explained at least partially by the association between child-care quality and academic skills at entry to school, which was then maintained into high school. This mediated pathway accounts for 47% of the total association between quality and age 15 cognitive–academic achievement and 29% of the total association between quality-squared and age 15 cognitive–academic achievement.

A second pathway model tested the extent to which child-care hours was related to youth reports of behavior problems, risk taking, and impulsivity at age 15 though the maintenance of earlier teacher-reported externalizing behaviors. The model fit these data adequately well despite the absence of latent variables in the model (RMSEA = .09). Results are shown in Figure 5.

Significant direct paths remained in these analyses from child-care hours to age 15 assessments of risk taking ($B = 0.007$, $SE = 0.003$, $p < .05$, $d = .08$)

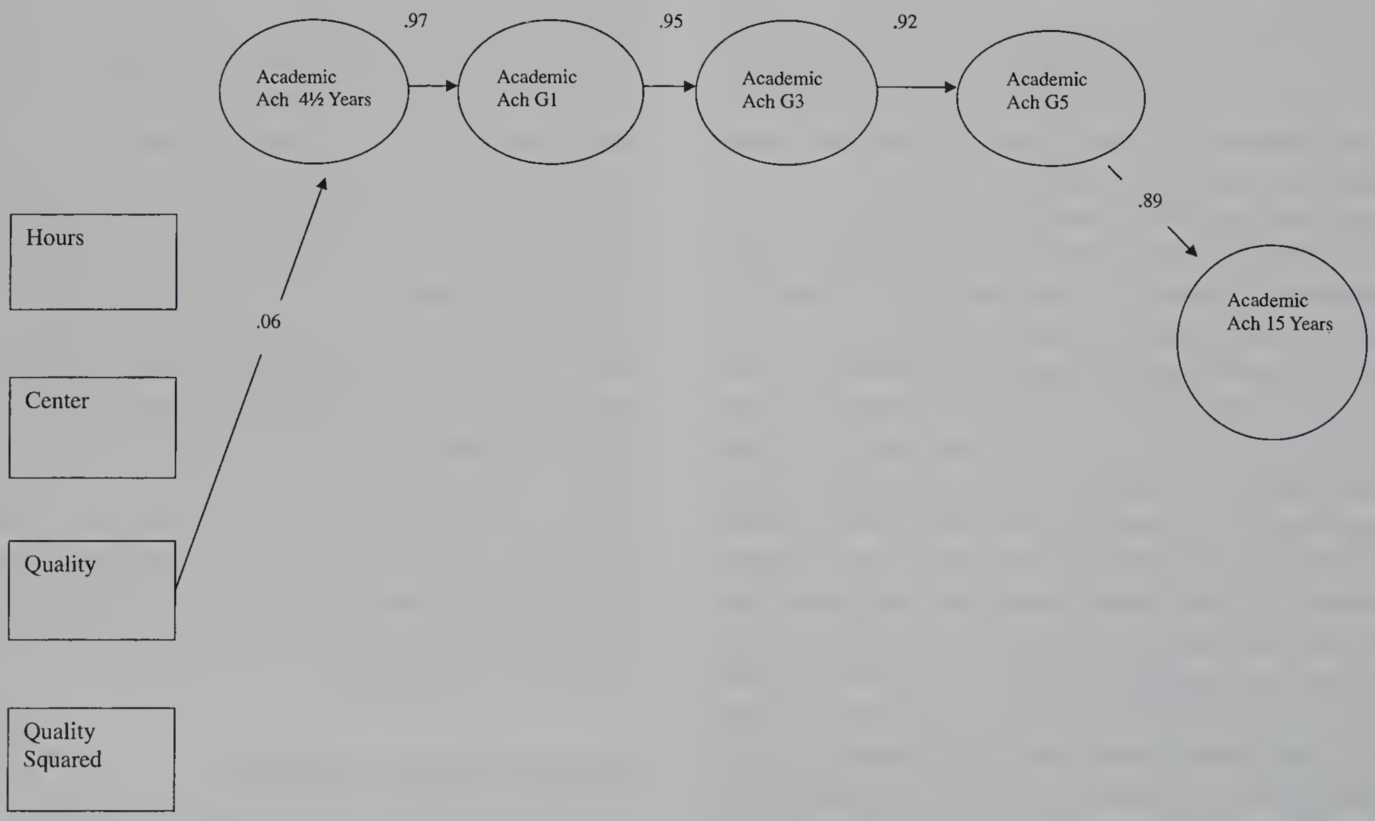


Figure 4. Pathway structural equation model testing of prior academic–cognitive achievement as a mediator of the link between child-care quality and age 15 academic–cognitive achievement.

Note. Only significant paths are depicted. Covariates were site, gender, ethnicity, maternal education, maternal psychological adjustment, maternal Peabody Picture Vocabulary Test–Revised, elementary school classroom quality and repeated assessments of family income, proportion time mother had a husband/partner, maternal depressive symptoms, and parenting. Ach = achievement; G1 = Grade 1; G3 = Grade 3; G5 = Grade 5.

and impulsivity ($B = 0.07$, $SE = 0.03$, $p < .001$, $d = .12$) and from child-care quality to externalizing ($B = -1.82$, $SE = 0.86$, $p < .05$, $d = -.08$), remained, suggesting that the hypothesized mediators—earlier caregiver- or teacher-reported externalizing problems—did not account for most of the observed effect of hours on these adolescent behavioral outcomes although the mediated paths from child-care hours through teacher ratings of externalizing behavior between preschool and Grade 5 were significant for both risk taking ($B = 0.001$, $SE = 0.001$, $p < .01$, $d = .003$) and impulsivity ($B = 0.002$, $SE = 0.001$, $p < .01$, $d = .003$). As a percentage of the total effect, these mediated effects were small: 4% of the total path to risk taking and 3% of total path to impulsivity.

Early Risk and Gender as Moderators

The final set of analyses addressed questions of moderation. One multiple group analysis tested familial risk and a second multiple group analysis tested gender as moderators of early child-care effects.

A cumulative familial risk index was created as the standardized mean of maternal education (reflected), mean income-to-needs ratio from 6 to 54 months (reflected), proportion time with single parent from 6 to 54 months, mean HOME total score from 6 to 54 months (reflected), and mean maternal sensitivity from 6 to 54 months (reflected). Families in the top quartile (high risk), middle 50%, and bottom quartile were compared with test hypotheses about compensatory effects and lost resources. Families in high-risk group had, on average, mothers with less than a high school education, an income-to-needs ratio of less than 1.3, CES–D depressive symptoms of 15.5, and a husband or partner less than 50% of the time. The low-risk group had, on average, mothers with a BA or more, income-to-needs ratio of 6.7, a husband or partner 99% of the time, and CED–D depressive symptom scores of 5.1.

A multigroup SEM allowed for different paths between child-care experiences and 15-year outcomes for the three risk groups. Due to the complexity of the multiple group SEMs for the three risk groups, separate models tested risk as a moderator for the behavioral outcomes and for the cognitive–achievement outcomes. The high correlation between hours of nonrelative care and time in center care prevented the multigroup

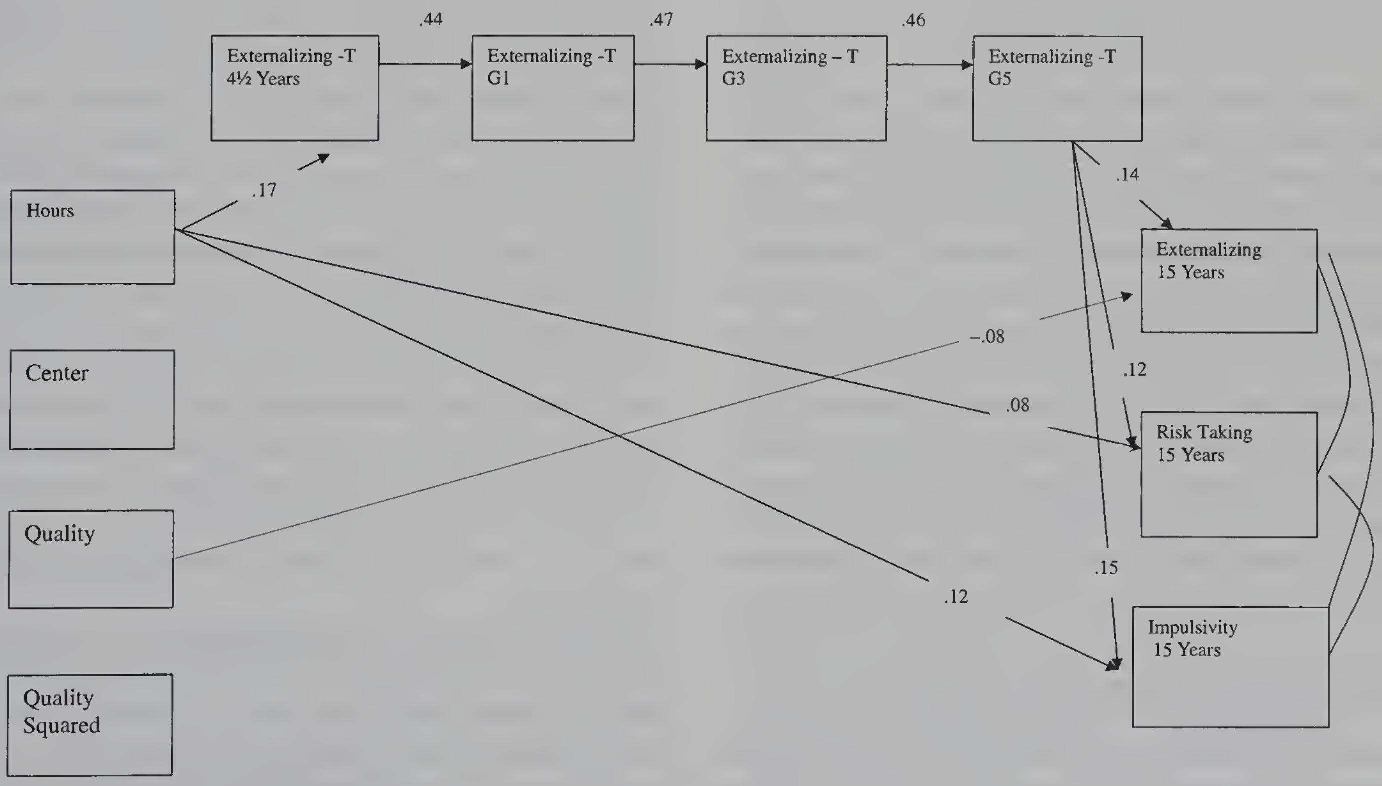


Figure 5. Pathway structural equation model testing teacher ratings of externalizing at earlier periods as a mediator of the link between child-care hours and quality and problem behaviors at age 15.

Note. Only significant paths are depicted. Covariates include site, gender, ethnicity, maternal education, maternal Peabody Picture Vocabulary Test–Revised, maternal psychological adjustment, elementary school classroom quality and repeated assessments of family income, proportion of time mother had a husband or partner, maternal depressive symptoms, and parenting quality. T = teacher report; G1 = Grade 1; G3 = Grade 3; G5 = Grade 5.

analyses from converging; hence, separate models were run to look at hours and center care as moderators.

The overall likelihood-ratio test between the model with and without separate paths from child-care experiences to behavioral outcomes was not reliably different among the three risk groups for the model that included hours and quality, $\chi^2(724) = 28.72$, $p = .99$, or for the model that included center care and quality, $\chi^2(838) = 58.26$, $p = .14$, suggesting no support for the compensatory hypothesis (i.e., stronger positive paths for high-risk children) or the lost resources hypothesis (i.e., stronger negative paths for low-risk children). Similarly, no evidence emerged in the analysis of the cognitive–academic achievement outcomes supporting either the compensatory or lost resources hypotheses.

A similar strategy was used to test whether gender moderated associations between the child-care variables and age 15 outcomes. Again, the likelihood-ratio test between the model with and without gender-specific paths from child-care experiences to child outcomes was nonsignificant, likelihood-ratio test $\chi^2(206) = 243.8$, $p = .13$, suggesting that gender did not moderate these associations.

Discussion

This latest installment in a 15-year longitudinal study of the effects of early child care on academic and behavioral development addressed three questions: (a) Are early child-care quality, quantity, and type related to adolescent functioning (cognitive–academic achievement and behavior problems) at age 15? (b) Are pathways from early child care to adolescent functioning mediated through prior functioning? and (c) Are relations between early child care and adolescent development moderated by either child gender or familial risk? Results pertaining to cognitive–academic achievement at age 15 are discussed first, followed by those pertaining to behavior problems.

Cognitive–Academic Achievement at Age 15

A relatively consistent finding across 40 years of child-care research is that quality and type of care are related to cognitive, academic, and language functioning in young children (Belsky & Steinberg, 1978; Lamb & Ahnert, 2006; Vandell, 2004). Previous findings from the NICHD SECCYD are consistent with this conclusion. At 4½ years, higher quality care predicted higher levels of preacademic skills ($d = .16$) and language ($d = .10$), whereas more exposure to center-type care predicted better language ($d = .11$) and memory ($d = .11$; NICHD ECCRN, 2002). In this report of adolescent functioning measured more than 10 years after the children had left child care, we find early child-care quality continues to predict cognitive–academic achievement. The size of the effect ($d = .09$) is similar in magnitude to that detected at 4½ years.

This evidence of long-term effects of early child-care quality is noteworthy because it occurred in a large economically and geographically diverse group of children who participated in routine non-relative child care in their communities. Previous long-term studies have focused on high-quality interventions aimed at children at risk because of poverty or low birth weight. The current findings suggest that the quality of early child-care experiences can have long-lasting (albeit small) effects on middle class and affluent children as well as those who are economically disadvantaged.

In addition to a linear relation between child-care quality and adolescent cognition, we detected a significant quadratic relation at age 15 (effect size = .07). This quadratic effect indicated that child-care quality was linked to academic outcomes for those adolescents whose care, on average was of moderate quality or better, with the magnitude of the quality effects being larger at higher levels of quality. Intriguingly, similar findings have been reported in an 11-state PreK evaluation that examined concurrent associations between child-care quality and academic outcomes (Burchinal et al., 2009). That study also reported escalating effect sizes linked to quality in the moderate to high quality range.

These two studies, taken together, underscore two points: that larger gains in cognitive–academic outcomes appear to accrue when children experience care of high quality, and that improvements in child-care quality in the moderate to high range may be needed to yield measurable long-term benefits.

This report is the first paper from the SECCYD data set to report this nonlinear quality effect, which raises the question of whether the effect was present at earlier ages but not detected because of different analytic strategies or whether this is a "new" finding (a sleeper effect). To address the first possibility, we examined relations between the quadratic term and earlier cognitive–academic functioning using the same models and operational variables as this study. We found no evidence of a quadratic relation at earlier ages. Therefore, the quadratic relation may be viewed as a sleeper effect. It remains to be seen whether this nonlinear effect carries forward to late adolescence and early adulthood or is found for other developmental domains.

Additional research is needed to better understand *why* the longer term effects in adolescence were evident only when early quality was in the moderate to high range. One possibility is that high school students bear much greater responsibility for their own academic progress whereas instruction for younger students is more closely monitored and supervised. Students who had attended higher quality programs (and who had excelled cognitively and academically) may be better positioned to oversee their own achievement in high school.

Mediated Pathways

An additional issue examined in this report was the mechanism or process that might mediate the long-term relation between early child care and age 15 functioning. It was anticipated that the cognitive–academic benefits of child-care quality observed at 4½ years would be carried forward to age 15 through functioning during middle childhood. As hypothesized, this mediated pathway was significant, accounting for 47% of the linear path and 29% of the quadratic path from child-care quality to adolescent cognitive functioning. These results are consistent with theory and empirical evidence that emphasize developmental continuities whereby competencies in one period set the stage for and are then carried forward to later periods (Bronfenbrenner & Morris, 2006; Burchinal, Peisner-Feinberg, Pianta, & Howes, 2002; Campbell et al., 2001). Because school achievement is largely cumulative, it is not a surprise to find that higher achievement in adolescence is linked to higher achievement in earlier periods.

Behavior Problems

Another consistent finding in the literature is that more hours in child care and more center-type care are related to higher levels of behavior problems in young children (Belsky, 2001; Lamb & Ahnert, 2006; Loeb et al., 2007). Previous findings from the SECCYD support this observation. At 4½ years of age, caregivers reported higher levels of externalizing behaviors for children who were in care for more hours ($d = .16$) and who had more center-type experience ($d = .11$; NICHD ECCRN, 2002). In this article, surprisingly similar relations were detected between child-care hours and problem behaviors at age 15. Higher hours predicted reports by the adolescents of more risk taking ($d = .09$) and greater impulsivity ($d = .13$). These effect sizes are similar in magnitude to the relation between hours and caregivers' reports of externalizing behavior problems originally detected at 4½ years.

We then sought to determine the extent to which associations between child-care hours and age 15 problem behaviors were mediated by behavior problems at 4½ years that were carried forward. Although we found statistically significant evidence of mediation, this mediated path accounted for only 3%–4% of the effect of child-care hours on adolescent problem behaviors. This modest level of mediation is likely because of two factors. The informant who reported the behavior problems changed from the caregivers in early childhood and teachers in middle childhood to the adolescent in this report. In addition, two of the three measures at 15 years assessed somewhat different aspects of behavior problems, ones that are particularly pertinent to adolescence: risk taking and impulsivity. Thus, although there is a large literature documenting continuity in behavior problems between childhood and adolescence (Farrington, 2004), we suspect that the very modest levels of mediation detected in this analysis are because of these changes in measurement.

Our age 15 analyses also revealed a relation between higher quality nonrelative child care and less externalizing behavior. This relation harkens back to findings detected when children were toddlers, when higher quality child care predicted fewer behavior problems at 2 and 3 years of age according to caregivers (NICHD ECCRN, 1998). We did not, however, find similar associations at 4½ years or middle childhood (Belsky et al., 2007; NICHD ECCRN, 2002, 2003, 2005c). Because the current analyses varied in a number of ways from those conducted at earlier ages, follow-up analyses examined externalizing at 4½ years of age and middle childhood using the same SEM models conducted for all analyses in this article. Some modest associations were found in these reanalyses, indicating that the link between higher quality of nonrelative care and fewer externalizing problems may not be as "new" as it might otherwise appear.

Small Enduring Effects

Although the obtained child-care effects on cognitive–academic outcomes and problem behaviors are small by conventional standards, we would argue that they should not be dismissed. First, they endured over a 10-year period at roughly the same size, suggesting a consistent pattern of relations. Second, as observed by McCartney and Rosenthal (2000), it is useful to benchmark effects against those reported in other studies that have been judged to be important or meaningful. A useful comparison to this study is the 18-year follow-up of the participants in the Infant Health and Development Program (IHDP; McCormick et al., 2006), an intensive early intervention program for low-birth-weight children that included home visits in the first 3 years and high-quality center care for 1–3 years. The IHDP reported significant effects for the heavier babies: effect sizes of .34 on math achievement and .27 for (reduced) risky behaviors at the 18-year follow-up, and no effects for the smallest of the low-birth-weight babies. The long-term effects associated with quality and quantity of child care in the SECCYD after 10 years are 29%–44%, respectively, as large as those found for the heavier infants in this *intensive* experimental intervention.

A third reason not to discount the findings is that child care, often full-time care beginning very early in life, has become a normative experience for American children, meaning that the number of children directly affected by child care is large. Small effects distributed over many people may have cumulative influences. Indeed, recent evidence suggests that children without child-care experience may be influenced by their classmates with early child care. Children in kindergarten classrooms in which there were higher proportions of children with child-care experience evinced better academic achievement but also more behavior problems than children in less child-care-saturated classrooms, even when the children themselves had not attended child care (Dmitrieva, Steinberg, & Belsky, 2007). Future research should more carefully examine such peer effects at older ages, especially as they may have implications for detecting child-care effects over the longer term (Belsky, 2009).

No Evidence of Moderation

Although two recent long-term follow-ups of early educational interventions have reported gender-moderated effects (Anderson, 2008; Belfield et al., 2006), we found no evidence of differential child-care effects as a function of child gender in either cognitive–academic performance or problem behaviors. This result is entirely consistent with findings discerned at earlier ages. Because our sample size is large and there is good variability in the measures of child-care experience and child functioning, it seems unlikely that the absence of significant gender-moderated child-care effects can be attributed to lack of statistical power. Perhaps secular changes in the 1990s, when maternal employment and nonmaternal child care became normative (i.e., characteristic of the majority of households in the United States), contributed to the similar developmental pathways among adolescent boys and girls observed in this study.

Evidence that high-quality care and/or center care are especially beneficial for economically disadvantaged children has emerged in a number of studies (Gormley et al., 2005; Loeb et al., 2007; Magnuson et al., 2007). We, however, did not detect any evidence that the effects of early child care (in quality, hours, type) on adolescent outcomes were moderated by familial risk, a pattern consistent with previous reports of the SECCYD (Belsky et al., 2007; NICHD ECRN, 2002, 2005c), although our high- and low-risk groups were decidedly different from one another. However, the exclusion of adolescent mothers and non-English speakers from the study may have limited our power to evaluate adequately this differential effect.

No Evidence of Long-Term Effects of Center Care

Robust center care effects have been chronicled in several large nationally representative surveys (Loeb et al., 2007; Magnuson et al., 2007), as well as evaluations of state prekindergarten initiatives (Gormley et al., 2005). Previous analyses of the SECCYD also have detected center care effects on both cognitive and social outcomes at every age through sixth grade. One of the most surprising findings of the current analyses was the failure to detect reliable associations between the proportion of time spent in center care and the age 15 outcomes explored in this report. One possible explanation is that children who experienced longer hours tended to have more center care ($r = .51$), thereby masking unique center care effects. However, this explanation was ruled out when follow-up analyses failed to discern center care effects even when hours were excluded from the model. Interestingly, another recent analysis based on the SECCYD did detect relations between center-type care and other age 15 outcomes—more center-type care experience was linked to blunted awakening cortisol levels (Roisman et al., 2009)—so it would seem misguided to conclude broadly that by mid-adolescence type of care no longer predicts children's development. As we continue to follow the sample, it will be important to determine if center care effects on cognitive and social functioning reemerge in late adolescence or early adulthood.

Finally, there are limitations to the study that should be noted. First, the study design is correlational, not experimental. So, analyses were tests of associations, not causation. Although extensive covariates were included in the analyses, omitted variables may account for the obtained effects. Second, although the sample was diverse economically and geographically, the study sample was not constructed to be nationally representative.

Conclusion

Perhaps the most important findings of this report from the SECCYD is that the effects of early child-care quality on cognitive–academic achievement and early child-care hours on problem behaviors were evident in mid-adolescence, more than a decade after the children had transitioned from child care to elementary school. Effect sizes were small but comparable in size with those detected at earlier ages. These findings extend research from other projects that documented the impact of high-quality child-care interventions for economically disadvantaged children into adolescence and adulthood. This study is the first to document relations between routine nonrelative child care and adolescent functioning for children from economically diverse families.

It remains to be determined whether, as individuals develop from adolescence into adulthood, the apparent consequences of child care are sustained, dissipate, or increase. To the extent that early child-care quality increases cognitive–academic skills, it will be important to learn whether subsequent educational attainment in high school and beyond is related to early child-care experience. To the extent that early child care increases adolescent impulsivity and risk taking, it is important to follow individuals into late adolescence, when opportunities for engaging in risky behavior increase and the capacity for self-regulation is still not fully mature.

References

Achenbach, T. M. (1991a). *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. Burlington: University of Vermont.

Achenbach, T. M. (1991b). *Manual for the Teacher's Report Form and 1991 Profile*. Burlington: University of Vermont.

Achenbach, T. M., & Rescorla, L. (2001). *Manual for the ASEBA (Achenbach System of Empirically-Based Assessment) School-Age Forms and Profiles*. Burlington: University of Vermont, Research Center for Children, Youth, and Families.

Allen, J. P., Hall, F. D., Insabella, G. M., Land, D. J., Marsh, P. A., & Porter, M. R. (2001). *Supportive behavior coding system*. Charlottesville: University of Virginia.

Allen, J. P., Hauser, S. T., Bell, K. L., McElhaney, K. B., Tate, D. C., Insabella, G. M., et al. (2000). *The autonomy and relatedness coding system*. Unpublished manuscript, University of Virginia, Charlottesville.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association, 103*, 1481–1495.

Bates, J. E., Marvinney, D., Kelly, T., Dodge, K. A., Bennett, D. S., & Pettit, G. S. (1994). Child-care history and kindergarten adjustment. *Developmental Psychology, 30*, 690–700.

Belfield, C. R., Nores, M., Barnett, S., & Schweinhart, L. (2006). The High/Scope Perry Preschool Program: Cost benefit analysis using data from the age-40 follow up. *Journal of Human Resources, 61*, 162–190.

Belsky, J. (1986). Infant day care: A cause for concern? *Zero to Three, 6*, 1–7.

Belsky, J. (1988). The "effects" of infant day care reconsidered. *Early Childhood Research Quarterly, 3*, 235–272.

Belsky, J. (2001). Developmental risks (still) associated with early child care. *Journal of Child Psychology and Psychiatry, 42*, 845–859.

Belsky, J. (2009). Classroom composition, child-care history and social development: Are child-care effects disappearing or spreading? *Social Development, 18*, 230–238.

Belsky, J., & Steinberg, L. D. (1978). The effects of day care: A critical review. *Child Development, 49*, 929–949.

Belsky, J., Vandell, D., Burchinal, M., Clarke-Stewart, K. A., McCartney, K., Owen, M., et al. (2007). Are there long-term effects of early child care? *Child Development, 78*, 681–701.

Blau, D. (1999). The effects of child care characteristics on child development. *Journal of Human Resources, 34*, 786–822.

Bollen, K. A. (1989). *Structural equations with latent variables. Wiley series in probability and mathematical statistics*. New York: Wiley.

Bradley, R. H., Caldwell, B. M., & Rock, S. L. (1988). Home environment and school performance: A ten-year follow-up and examination of three models of environmental action. *Child Development, 59*, 852–867.

Bradley, R. H., Corwyn, R. F., Caldwell, B. M., Whiteside-Mansell, L., Wasserman, G. A., & Mink, I. T. (2000). Measuring the home environments of children in early adolescence. *Journal of Research on Adolescence, 10*, 247–288.

Broberg, A. G., Wessels, H., Lamb, M. E., & Hwang, C. P. (1997). Effects of day care on the development of cognitive abilities in 8-year-olds: A longitudinal study. *Developmental Psychology, 33*, 62–69.

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In W. Damon & R. M. Lerner (Series Eds.) & R. M. Lerner (Vol. Ed.), *Handbook of child psychology: Vol. 1. Theoretical models of human development* (6th ed., pp. 793–828). Hoboken, NJ: Wiley.

Burchinal, M. R., Peisner-Feinberg, E., Pianta, R., & Howes, C. (2002). Development of academic skills from preschool through second grade: Family and classroom predictors of developmental trajectories. *Journal of School Psychology, 40*, 415–436.

Burchinal, M. R., Roberts, J. E., Riggins, R., Zeisel, S., Neebe, E., & Bryant, M. (2000). Relating quality of center child care to early cognitive and language development longitudinally. *Child Development, 71*, 339–357.

Burchinal, M. R., Vandergrift, N., & Pianta, R. (2009). *Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs*. Manuscript under review.

Campbell, F. A., Pungello, E. P., Miller-Johnson, S., Burchinal, M., & Ramey, C. T. (2001). The development of cognitive and academic abilities: Growth curves from an early childhood educational experiment. *Developmental Psychology, 37*, 231–242.

Caughy, M. O., DiPietro, J. A., & Strobino, D. M. (1994). Day-care participation as a protective factor in the cognitive development of low-income children. *Child Development, 65*, 457–471.

Collins, W. A., & Steinberg, L. (2006). Adolescent development in interpersonal context. In W. Damon & R. M. Lerner (Series Ed.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (6th ed., pp. 1003–1067). Hoboken, NJ: Wiley.

Colwell, M., Pettit, G., Meece, D., Bates, J. E., & Dodge, K. A. (2001). Cumulative risk and continuity in nonparental care from infancy to early adolescence. *Merrill-Palmer Quarterly, 47*, 207–234.

Committee on Family and Work Policies. (2003). *Working families and growing kids: Caring for children and adolescents*. Washington, DC: National Academies Press.

Costa, P., & McCrae, R. (1985). *NEO Personality Inventory Manual*. Odessa, FL: Psychological Assessment Resources.

Costa, P., & McCrae, R. (1989). *NEO Five-Factor Inventory (NEO-FFI)*. Odessa, FL: Psychological Assessment Resources.

Cote, S. M., Boivin, M., Nagin, D. S., Japel, C., Xu, Q., Zoccolillo, M., et al. (2007). The role of maternal education and nonmaternal care services in the prevention of children's physical aggression problems. *Archives of General Psychiatry, 64*, 1305–1312.

Cote, S. M., Borge, A. I., Geoffroy, M., Rutter, M., & Tremblay, R. E. (2008). Nonmaternal care in infancy and emotional/behavioral difficulties at 4 years old: Moderation by family risk characteristics. *Developmental Psychology, 44*, 155–168.

Crockenberg, S. C. (2003). Rescuing the baby from the bathwater: How gender and temperament may influence how child care affects child development. *Child Development, 74*, 1034–1038.

Darlington, R., Murray, H., Royce, J., Snipper, A., & Ramey, C. T. (1982). Lasting effects of early education. *Monographs of the Society for Research in Child Development, 47* (Serial No. 195).

Deater-Deckard, K. A., Pinkerton, R., & Scarr, S. (1996). Child care quality and children's behavioral adjustment: A four-year longitudinal study. *Journal of Child Psychology and Psychiatry, 37*, 937–948.

Desai, S., Chase-Lansdale, P. L., & Michael, R. T. (1989). Mother or market? Effects of maternal employment on the intellectual ability of 4-year-old children. *Demography, 26*, 545–561.

Dmitrieva, J., Steinberg, L., & Belsky, J. (2007). Child-care history, classroom composition, and children's functioning in kindergarten. *Psychological Science, 18*, 1032–1039.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Test–Revised manual for Forms L and M*. Circle Pines, MN: American Guidance Service.

Eccles, J., & Roeser, R. (2009). Schools, academic motivation, and stage-environment fit. In R. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology* (3rd ed., pp. 404–434). Hoboken, NJ: Wiley.

Egeland, B., & Hiester, M. (1995). The long-term consequences of infant day-care and mother–infant attachment. *Child Development, 66*, 474–485.

Farrington, D. (2004). Conduct disorder, aggression, and delinquency. In R. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology* (2nd ed., pp. 627–664). Hoboken, NJ: Wiley.

Gormley, W. T., Jr., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal pre-K on cognitive development. *Developmental Psychology, 41*, 872–884.

Gottfried, A. E., & Gottfried, A. W. (1988). *Maternal employment and children's development*. New York: Plenum.

Halpern-Felsher, B. L, Biehl, M., Kropp, R. Y., & Rubinstein, M. L. (2004). Perceived risks and benefits of smoking: Differences between adolescents with different smoking experiences and intentions. *Preventive Medicine, 39*, 559–567.

Haskins, R. (1985). Public school aggression among children with varying day-care experience. *Child Development, 56*, 689–703.

Hoffman, L. W., & Youngblade, L. M. (1999). *Mothers at work: Effects on children's well-being*. New York: Cambridge University Press.

Howes, C., Phillips, D. A., & Whitebook, M. (1992). Thresholds of quality: Implications for the social development of children in center-based child care. *Child Development*, *63*, 449–460.

Huston, A. C., Duncan, G. J., Granger, R., Bos, J., McLoyd, V., Mistry, R., et al. (2001). Work-based antipoverty programs for parents can enhance the school performance and social behavior of children. *Child Development*, *72*, 318–336.

Kovan, N. M., Chung, A. L., & Sroufe, L. A. (2009). The intergenerational continuity of observed early parenting. *Developmental Psychology*, *45*, 1205–1213.

Kuhn, D. (2009). Adolescent thinking. In R. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology* (3rd ed., pp. 152–186). Hoboken, NJ: Wiley.

Lamb, M. E., & Ahnert, L. (2006). Nonparental child care: Context, concepts, correlates, and consequences. In W. Damon & R. M. Lerner (Series Eds.) & K. A. Renninger, & I. E. Sigel (Vol. Eds.), *Handbook of child psychology: Vol. 4. Child psychology in practice* (6th ed., pp. 950–1016). Hoboken, NJ: Wiley.

Lamb, M. E., Thompson, R. M., & Gardner, W. (1984). Security of infantile attachment as assessed in the strange situation. *Behavioral and Brain Sciences*, *7*, 127–147.

Loeb, S., Bridges, M., Bassok, D., Fuller, B., & Rumberger, R. W. (2007). How much is too much? The influence of preschool centers on children's social and cognitive development. *Economics of Education Review*, *26*, 52–66.

Love, J. M., Harrison, L., Sagi-Schwartz, A., van IJzendoorn, M. H., Ross, C., Ungerer, J. A., et al. (2003). Child care quality matters: How conclusions may vary with context. *Child Development*, *74*, 1021–1033.

Love, J. M., Kisker, E. E., Ross, C., Raikes, H., Constantine, J., Boller, K., et al. (2005). The effectiveness of Early Head Start for 3-year-old children and their parents. *Developmental Psychology*, *41*, 885–901.

Maccoby, E. E., & Lewis, C. C. (2003). Less day care or different day care? *Child Development*, *74*, 1069–1075.

Magnuson, K. A., Ruhm, C., & Waldfogel, J. (2007). Does prekindergarten improve school preparation and performance? *Economics of Education Review*, *26*, 33–51.

Mashburn, A. J., Pianta, R. C., Barbarin, O., Bryant, D., Hamre, B., Downer, J., et al. (2008). Measures of classroom quality in pre-kindergarten and children's development of academic, language and social skills. *Child Development*, *79*, 732–749.

McCartney, K., & Rosenthal, R. (2000). Effect size, practical importance, and social policy for children. *Child Development*, *71*, 173–180.

McCormick, M. C., Brooks-Gunn, J., Buka, S. L., Goldman, J., Yu, U., Salganik, M., et al. (2006). Early intervention in low birthweight premature infants: Results at 18 years of age for the Infant Health and Development Program (IHDP). *Pediatrics*, *117*, 771–780.

Melhuish, E. C., Sylva, K., Sammons, P., Siraj-Blatchford, I., Taggart, B., Phan, M. B., et al. (2008). Preschool influences on mathematics achievement. *Science*, *321*, 1161–1162.

NICHD Early Child Care Research Network. (1998). Early child care and self-control, compliance, and problem behavior at twenty-four and thirty-six months. *Child Development*, *69*, 1145–1170.

NICHD Early Child Care Research Network. (2002). Early child care and children's development prior to school entry: Results from the NICHD Study of Early Child Care. *American Education Research Journal*, *39*, 133–164.

NICHD Early Child Care Research Network. (2003). Does amount of time spent in child care predict socioemotional adjustment during the transition to kindergarten? *Child Development*, *74*, 976–1005.

NICHD Early Child Care Research Network. (2004). Does class size in first grade relate to children's academic and social performance or observed classroom processes? *Developmental Psychology*, *40*, 651–664.

NICHD Early Child Care Research Network. (2005a). A day in third grade: A large-scale study of classroom quality and teacher and student behavior. *Elementary School Journal*, *105*(3), 305–323.

NICHD Early Child Care Research Network. (2005b). *Child care and child development: Results of the NICHD Study of Early Child Care and Youth Development*. New York: Guilford.

NICHD Early Child Care Research Network. (2005c). Early child care and children's development in the primary grades: Results from the NICHD Study of Early Child Care. *American Educational Research Journal*, *43*, 537–570.

NICHD Early Child Care Research Network. (2008). Mothers' and fathers' support for child autonomy and early school achievement. *Developmental Psychology*, *44*, 895–907.

NICHD Early Child Care Research Network, & Duncan, G. J. (2003). Modeling the impacts of child care quality on children's preschool cognitive development. *Child Development*, *74*, 1454–1475.

Nomaguchi, K. M. (2006). Maternal employment, nonparental care, mother–child interactions, and child outcomes during preschool years. *Journal of Marriage and the Family*, *68*, 1341–1369.

Peisner-Feinberg, E., & Burchinal, M. (1997). Concurrent relations between child care quality and child outcomes: The study of cost, quality, and outcomes in child care centers. *Merrill-Palmer Quarterly*, *43*, 451–477.

Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., & Kagan, S. L. (2001). The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second grade. *Child Development*, *72*, 1534–1553.

Radloff, L. S. (1977). The CES–D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*, 385–401.

Reynolds, A. J. (2000). *Success in early childhood interventions: The Chicago Child–Parent Centers*. Lincoln: University of Nebraska.

Roisman, G. I., Susman, E., Barnett-Walker, K., Booth-LaForce, C., Owen, M. T., Belsky, J., et al. (2009). Early family and child-care antecedents of awakening cortisol levels in adolescence. *Child Development*, *80*, 907–920.

Schafer, J., & Graham, J. (2002). Missing data: Our view on the state of the art. *Psychological Methods*, *7*, 147–177.

Schweinhart, L. J., Weikart, D. P., & Larner, M. B. (1986). Consequences of three preschool curriculum models through age 15. *Early Childhood Research Quarterly*, *1*, 15–45.

Susman, E., & Dorn, L. (2009). Puberty: Its role in development. In R. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology* (3rd ed., pp. 116–151). New York: Wiley.

Vandell, D. L. (2004). Early child care: The known and the unknown. *Merrill-Palmer Quarterly*, *50*, 387–414.

Vandell, D. L., & Corasaniti, M. A. (1990). Variations in early child care: Do they predict subsequent social, emotional, and cognitive differences. *Early Childhood Research Quarterly*, *5*, 555–572.

Vandell, D. L., Henderson, L.V., & Wilson, K. S. (1988). A longitudinal study of children with day-care experiences of varying quality. *Child Development*, *59*, 1286–1292.

Weinberger, D. A., & Schwartz, G.E. (1990). Distress and restraint as superordinate dimensions of self-reported adjustment: A typological perspective. *Journal of Personality*, *58*, 381–417.

Woodcock, R. W., & Johnson, M. B. (1989). *WJ–R Tests of Cognitive Ability*. Itasca, IL: Riverside.

Child Development, May/June 2010, Volume 81, Number 3, Pages 757–777

The Development of Social Essentialism: The Case of Israeli Children's Inferences About Jews and Arabs

Dana Birnbaum, Inas Deeb, and Gili Segall
Bar-Ilan University

Adar Ben-Eliyahu
Duke University

Gil Diesendruck
Bar-Ilan University

Two studies examined the inductive potential of various social categories among 144 kindergarten, 2nd-, and 6th-grade Israeli children from 3 sectors: secular Jews, religious Jews, and Muslim Arabs. Study 1—wherein social categories were labeled—found that ethnic categories were the most inductively powerful, especially for religious Jewish children. Study 2—wherein no social category labels were provided—found no differences across sectors either in the inductive potential of ethnic categories or in children's capacity to visually recognize social categories. These results stress the importance of labels and cultural background in children's beliefs about social categories. The implications of these findings for accounts of the development of social essentialism are discussed.

For the past few years, a growing body of literature suggests that people across the world conceive of social categories in a way akin to how they conceive of living kinds. Specifically, people believe that, like categories of animals (Gelman, 2003), human categories (Gil-White, 2001) have unique properties that are innate and causal and that membership in these categories is absolute, permanent, and inductively powerful—a belief Medin and Ortony (1989) termed ''psychological essentialism.'' In pioneering work among North American children, Hirschfeld (1996) found that kindergarteners believe racial identity to be absolute, established at birth, and thus independent of upbringing environment (see also Taylor, 1996, on gender). Analogous findings have been reported on other cultures. For instance, Mahalingam (2003) showed that Indian Brahmin adults believe that a person's caste is inherited from birth parents, rather than acquired from adoptive parents, and Gil-White (2001) found a similar nativist construal of ethnic differences among Mongolian Torguuds (see also Astuti, Solomon, & Carey, 2004).

This cross-cultural and developmental consistency in children and adults' tendency to essentialize human kinds on the one hand, and the cultural specificity of the particular categories essentialized on the other, has led researchers to ponder about the *origins* of such beliefs (see, e.g., Atran, 1990; Hirschfeld, 1996; Sperber, 1996). Manifest in these accounts is the need to understand how social essentialism emerges and to what extent it is susceptible to cultural and linguistic variations. The present studies attempt to do exactly that. In particular, we investigate how *one* component of essentialism is manifested in children from various *subcultural* groups, toward the *same* social categories. As briefly reviewed above, the cross-cultural developmental literature on social essentialism consists of studies assessing different components of essentialism (e.g., induction vs. innateness), conducted on different cultural groups (e.g., the Vezo

---

This research was conducted with the support of Grant 621/05, from the Israel Science Foundation, to Gil Diesendruck. We thank the principals, teachers, parents, and especially children for their participation in the studies. We also thank the following research assistants for their help with data collection (in alphabetical order): Alaa Abu-Mouck, Huda Abu-Mouck, Rabea Alian, Huda Atamna, Hadas Detelkremer, Tali Krupkin, Michal Levinsky, Salam Mousa, Shira Nadal, Kiram Qaadan, Mona Sachneny, Haneen Wattad, and Einav Yavlon. This paper is dedicated to the memory of Shira Nadel.

Correspondence concerning this article should be addressed to Gil Diesendruck, Department of Psychology, Bar-Ilan University, Ramat-Gan 52900, Israel. Electronic mail may be sent to dieseng@mail.biu.ac.il.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8103-0008

of Madagascar vs. North Americans), and targeting different social categories (e.g., ethnicity vs. race). One of the main goals of the present studies was to attempt to minimize variance in these three dimensions, so as to more decisively assess the potential effect of cultural input on children's beliefs.

The component we chose to focus on is the inductive power of social categories, that is, the extent to which knowledge of the category to which an object belongs allows one to infer properties of the object. Gelman and colleagues have documented how from a young age children are sensitive to the unique inductive power of animal categories (see Gelman, 2003, for a review). In the domain of human categories, Heyman and Gelman (2000) found that children draw inferences based on characters' personality traits rather than their physical appearance, and Diesendruck and haLevi (2006) found that for Jewish Israeli adults and kindergarteners, ethnicity is the most inductively powerful social category. The present studies follow up on these latter findings.

For historical and political reasons, the category "Arab" for Jewish Israelis and the category "Jew" for Arab Israelis are extremely salient and important. Studies reveal that even 3-year-old Israeli Jewish children are somewhat familiar with the category "Arab" and that negative stereotypes regarding Arabs grow stronger from kindergarten to high school (Bar-Tal, 1996; Pitner, Astor, Benbenishty, Haj-Yahia, & Zeira, 2003; Slone, Tarrasch, & Hallis, 2000; Teichman, 2001). Bar-Tal and Teichman (2005) argue that a potential source for the development of negative stereotypes is the portrayal of Arabs and Jews in the media in general, and school books in particular (see also Al-Haj, 2002; Soen, 2002).

Taking into account the cultural and ideological heterogeneity of Israeli society, particularly with respect to issues pertaining to ethnicity, in the present studies we tested children from three different subcultural "sectors" within the Israeli society: secular Jews, modern religious (Zionist) Jews, and Muslim Arabs. Our selection of the three subgroups was motivated both by the diversity of ideological commitments vis-à-vis ethnicity typified by these groups, and by attention to the representativeness of these groups within the Israeli society. Specifically, secular Jews represent the majority group in Israel, modern religious Jews manifest different values for religious and often political reasons, and Muslim Arabs constitute the minority about whom the focal majority stereotypes are held (see the Participants section for further details). Evidence for cross-group differences would reinforce the notion that social essentialism is influenced by the type of input children are exposed to.

In order to trace the development of these potential cultural differences, we tested children from three age groups: kindergarteners, second graders, and sixth graders. Studies in North America found that these developmental points capture the transition from intuitive to more adult-like beliefs about human kinds (Hirschfeld, 1996; Taylor, 1996). Thus, in contrast to previous cross-cultural studies on social essentialism (Astuti et al., 2004; Gil-White, 2001; Mahalingam, 2003), the present design allowed us to assess how children's concepts change as they immerse into the institutionalized views exposed in schools.

Study 1 employed a methodology similar to the one used by Diesendruck and haLevi (2006). The experimenter presented children with pictures of two test characters, told children the category membership of these characters vis-à-vis two social categories (e.g., their ethnicity and gender), and informed children about a novel psychological property each test character had (e.g., their game preferences). The experimenter then presented to children a target character that resembled one of the test characters on one of the categories but resembled the other test character on the other category. Children were asked which of the two properties attributed to the test characters the target character possessed. The rationale was that children's pattern of responses across a number of trials would be indicative of the inductive power of the different social categories tested. To the extent that children's essentialist beliefs about certain social categories—especially ethnicity—is indeed susceptible to cultural input, we should find differences in the inductive power of these social categories across children from the different sectors. The age factor would allow us to assess how early these potential cultural differences appear.

A second major goal of the present studies was to continue investigating systematically the factors contributing to the inductive power of social categories, and their potential interaction with children's cultural group. A number of studies found that language (Bigler & Liben, 2007; Gelman, Taylor, & Nguyen, 2004; Heyman & Diesendruck, 2002), and perhaps especially labels (Cimpian, Arce, Markman, & Dweck, 2007; Gelman & Heyman, 1999; Patterson & Bigler, 2006), might be implicated in the development of social essentialism. The present studies investigate the role of language in children's essentialist reasoning by comparing children's pattern of induction when provided with category labels versus when only provided with physical appearance information about category membership. This comparison allowed us to verify to what extent appearance information about social category membership is sufficient to bring about potential differences among social categories. This examination is relevant to the broader question regarding whether children's essentialization of human kinds is a ''bottom-up'' process by which children simply represent categories they perceive in the world, or a ''top-down'' process by which the categories are first in the child's mind and then looked for in the world. The comparison is also relevant for understanding potential differences between cultural groups. Specifically, if a particular cultural group essentializes a certain social category, then will this bias be manifested only when labeling is present? In order to address these two related questions, Study 2 replicated part of Study 1, but whereas in Study 1 all the social categories were verbally labeled, in Study 2 none were.

Study 1

The goal of Study 1 was to assess the inductive power of various social categories across children from different cultural and age groups. For this purpose, we designed the induction task so as to both maximize the salience of social categories and allow a comprehensive assessment of their inductive potential. To these ends, the induction task included two parts. In the first—so-called relative power part—the two dimensions children had to decide between were both social categories. For instance, children had to decide whether to draw inferences based on the ethnic similarity between two characters, or their gender similarity. The maximization of the salience of social categories was operationalized by having all social categories represented by stereotypical visual markers associated with each social category, as well as their familiar category labels. The second—so-called absolute power part–was a direct replication of Diesendruck and haLevi (2006), and in it the two dimensions children had to decide between were personality traits (e.g., niceness) and social categories (e.g., ethnicity). This design feature augmented the comprehensiveness of our assessment, by widening the range of inferential contexts presented to children, including conflicts between different types of human categories. Moreover, Diesendruck and haLevi found that while kindergarteners had a preference for drawing inferences based on social categories, adults had a preference for drawing inferences based on personality traits. Thus, the inclusion of the absolute power part in the current study also allowed us to track in more detail the developmental changes in children's reliance on social categories versus personality traits as bases for inferences.

Method

Participants

A total of 144 children participated in this study. The sample consisted of 16 children from each of three sectors (secular Jews, religious Jews, and Muslim Arabs), and each of three age groups (kindergarten, second grade, and sixth grade). Only children with signed parental permission participated in the study. Table 1 presents age and gender information about the participants. The secular and religious Jewish children were recruited from secular and religious schools and kindergartens in the same cities surrounding the university. These cities have mixed populations of secular and religious Jews, though few Orthodox Jews, and practically no Arabs. The Arab children were recruited from an Arab and primarily Muslim city (Baka al Garbiah) within Israel. No Jews live in this city.

In general, the Israeli Jewish population is composed of three main religiosity subgroups: secular Jews (71%), religious Jews (21%), and Orthodox religious Jews (8%; Central Bureau of Statistics, 2006). We targeted secular and religious Jews for a variety of reasons. First, they represent the majority of the Jewish population. Second, both groups are socially, economically, and politically involved in Israeli life to similar extents, and to a greater extent than the Orthodox Jewish population. Third, despite these similarities, these two groups are denominationally distinguished, have substantially different belief systems and norms of behavior in everyday life, and are differentially distributed across the political spectrum (Sikron, 2004). Denominationally, Hebrew has two separate words, which are frequently used either as adjectives or count nouns, to refer to secular (*chiloni*) and religious (*dati*) Jews. Secular and religious groups diverge substantially in beliefs regarding God, the Jewish commandments, and afterlife, and in their practices, such as keeping the Sabbath and its commandments, eating Kosher food, celebrating the Jewish religious holidays, praying, and attending synagogues (Levy, Levinson, & Katz, 2002). Reflecting this divide, while most children from these two

Table 1
Participants' Age and Gender Distribution in Study 1 and Study 2

		Age group		
Study	Sector	Kindergarten	Second grade	Sixth grade
Study 1	Muslim Arabs	8 girls, 8 boys[a] M_{age} = 5 y 6 m (4 m, 4 y 10 m–6 y)	8 girls, 8 boys M_{age} = 7 y 6 m (5 m, 6 y 9 m–8 y 2 m)	13 girls, 3 boys M_{age} = 11 y 6 m (5 m, 10 y 9 m–12 y)
	Secular Jews	5 girls, 11 boys M_{age} = 5 y 5 m (4 m, 5 y 1 m–6 y)	10 girls, 6 boys M_{age} = 7 y 5 m (4 m, 7 y 1 m–8 y 6 m)	12 girls, 4 boys M_{age} = 11 y 7 m (6 m, 10 y 10 m–12 y 3 m)
	Religious Jews	9 girls, 7 boys M_{age} = 5 y 4 m (4 m, 5 y 1 m–6 y)	9 girls, 7 boys M_{age} = 7 y 6 m (4 m, 7 y 1 m–8 y 6 m)	14 girls, 2 boys M_{age} = 11 y 6 m (5 m, 11 y 1 m–12 y 2 m)
Study 2	Muslim Arabs	10 girls, 6 boys M_{age} = 5 y 5 m (4 m, 4 y 11 m–6 y 2 m)	8 girls, 8 boys M_{age} = 7 y 6 m (5 m, 6 y 9 m–8 y 2 m)	7 girls, 9 boys M_{age} = 11 y 7 m (5 m, 10 y 9 m–12 y 4 m)
	Secular Jews	6 girls, 10 boys M_{age} = 5 y 6 m (4 m, 4 y 11 m–6 y 2 m)	8 girls, 8 boys M_{age} = 7 y 10 m (4 m, 7 y 4 m–8 y 4 m)	13 girls, 3 boys M_{age} = 11 y 8 m (4 m, 10 y 10 m–12 y 3 m)
	Religious Jews	7 girls, 9 boys M_{age} = 5 y 5 m (6 m, 4 y 6 m–6 y)	13 girls, 3 boys M_{age} = 7 y 6 m (7 m, 7 y–8 y 6 m)	9 girls, 7 boys M_{age} = 11 y 10 m (5 m,11 y 2 m–12 y 5 m)

Note. y = years; m = months (*SD*, range).
[a] n = 16 per cell.

sectors attend the national public school system, they each attend the administratively and pedagogically separate branch that caters to each sector. In other words, the vast majority of students who attend secular schools come from secular families, and the vast majority of students who attend religious schools come from religious or traditional families. All major cities in Israel have public schools from each kind, and they are identifiable by their official names. One example of a major distinction between the two systems has to do with the extent of and approach toward teaching the Old Testament. In the secular system, the Old Testament is taught as a text recounting the *myths* and traditions of ancient Jews. In the religious system, the Old Testament is taught as a *veridical* account of the history of the world and the Jewish people, and the traditions are presented as normative rules to be followed presently as well. Lastly, these two sectors are characterized by distinct political orientations, especially toward the Israeli–Arab conflict. Namely, while the secular population endorses a variety of political opinions vis-à-vis the conflict, modern religious Zionist Jews tend toward opposing or stalling the creation of a Palestinian state in the West Bank and Gaza.

The Arab population in Israel is primarily Muslim (about 90%; Central Bureau of Statistics, 2006). In contrast to Hebrew, Palestinian Arabic does not have separate count nouns for ''religious'' and ''secular'' Muslims—all are referred to simply as ''Muslims.'' While there is variation in the degree of observance of religious practices, the majority (approximately 70%) of the Muslim population identifies itself as ''religious'' and rigorously observes the basic Islamic commandments, such as the belief in Allah and the prophet Mohamed, providing for the needy, praying five times a day, and fasting during the month of Ramadan. Moreover, less observant Arab Muslims nonetheless respect and attempt to incorporate some of these traditions into their daily lives (Stendel, 1996). Smooha (1989) observes that while since the establishment of the State of Israel in 1948, Arab Muslims have been undergoing a process of ''westernization,'' they still maintain their traditions and values, especially in cities and villages populated exclusively by Arab Muslims. The children recruited for the present study reside in such a city. Baka al Garbiah is located in an area of Israel referred to as the ''Triangle,'' is one of the largest Arab cities in Israel (over 30,000 inhabitants), and is almost exclusively Muslim. Over 90% of the children in this city attend public schools, which in contrast to the Jewish sector, all belong to the same pedagogical branch. These schools are administered by the Israeli

Ministry of Education, and they incorporate Islamic studies in their curricula. For instance, the schools celebrate all Muslim holidays, and there are at least 2 hr a week of classes on Islamic religion and history. Even though there are different Arab groups in Israel (e.g., Christians), for the sake of brevity, we refer to our Muslim Arab sample as "Arabs." In fact, while young Jewish Israeli children may be familiar with the term "Arab," it is unclear to what extent they differentiate between Muslims, Christians, Bedouins, or Druze.

Design

Study 1 consisted of two parts. In the first—"relative power"—part, children had to complete a series of 12 trials in which they had to draw inferences based on one of two possible social categories. This was named the relative power part because it assessed the inductive power of each social category relative to the other social categories. In the second—"absolute power"—part, children had to complete a series of 8 trials in which they had to draw inferences based either on a social category or a personality trait. This was named the absolute power part because in it, children did not have to choose between competing social categories but rather between each social category and a personality trait. The relative power part was always presented first, so as to allow an unbiased comparison between children's responses to this part across the two studies (to be described later).

The social categories used in both parts were: ethnicity (Jew/Arab), gender (boy/girl), social status (rich/poor), and religiosity (religious/secular). The four personality traits used in the absolute power part were "niceness" (nice/not nice), "shyness" (shy/friendly), "creativeness" (creative/not creative), and "activeness" (active/quiet). These categories and traits are salient in the Israeli society, and have been found to be familiar to Israeli kindergarten children (Diesendruck & haLevi, 2006).

Materials

In each trial, children were shown a triad of line-drawn black and white human characters: a target character and two test characters. Each character was professionally drawn with physical markers that are typical of specific social categories. In the relative power part, all three characters within a triad belonged to the same two social categories, but each was assigned a specific value of these categories (e.g., Arab or Jew for the category ethnicity). All possible combinations of social categories were presented to children, resulting in six types of triads: ethnicity/gender, ethnicity/social status, ethnicity/religiosity, gender/social status, gender/religiosity, and social status/religiosity. In each triad, each of the test characters resembled the target character on one of the social categories but was different from the target character on the other social category. For example, in a triad contrasting the social categories ethnicity and gender (as in Figure 1), one of the test characters was drawn as an Arabic boy, the other test character as a Jewish girl, and the target character as an Arabic girl. In this example, the test character "Arabic boy" resembled the target character ("Arabic girl") with regard to ethnicity (i.e., it was an "ethnicity match") but differed from her with regard to gender. In turn, the test character "Jewish girl" resembled the target character with regard to gender, but differed from her with regard to ethnicity (i.e., it was a "gender-match").

The visual cues consisted primarily of dressing attire and accessories typical of the social category. All Jewish characters had on their shirt or as a necklace the Star of David sign and a letter in Hebrew. Secular Arab characters wore on their shoulders Kaffiyah (a Palestinian scarf) and had a letter in Arabic on their shirt. Although there are no consistent visual cues that differentiate secular Arabs from secular Jews within Israeli society, the visual cues described above are typical markers of each ethnicity. In turn, there are distinct markers for both religious/practicing Jews and Muslims, and these were represented in the pictures. Specifically, male religious Jewish characters wore a yarmulke, fringes that come out of the shirt (*tzitziot*), and sidelocks (*peyot*). Female religious Jewish characters wore long dresses and had a head cover. Male religious Muslims wore a long robe and a skullcap. Female religious Muslims wore a head cover (*hijab*) and a long dress (*jilbab*). Rich characters wore fancy clothes and jewelry, and held expensive accessories. Poor characters wore ragged clothes and torn shoes or sandals. When the character was "religious" without a specific ethnicity, for example, a "religious boy" or "religious rich," it was depicted as a religious Jew for participants in the Jewish sector and as a religious Muslim for participants in the Arab sector. Figure 2 provides examples of the pictures used here.

In order to emphasize the relevant social categories in each triad in either part of the study, the physical characteristics of the characters not indicative of their social category membership were made

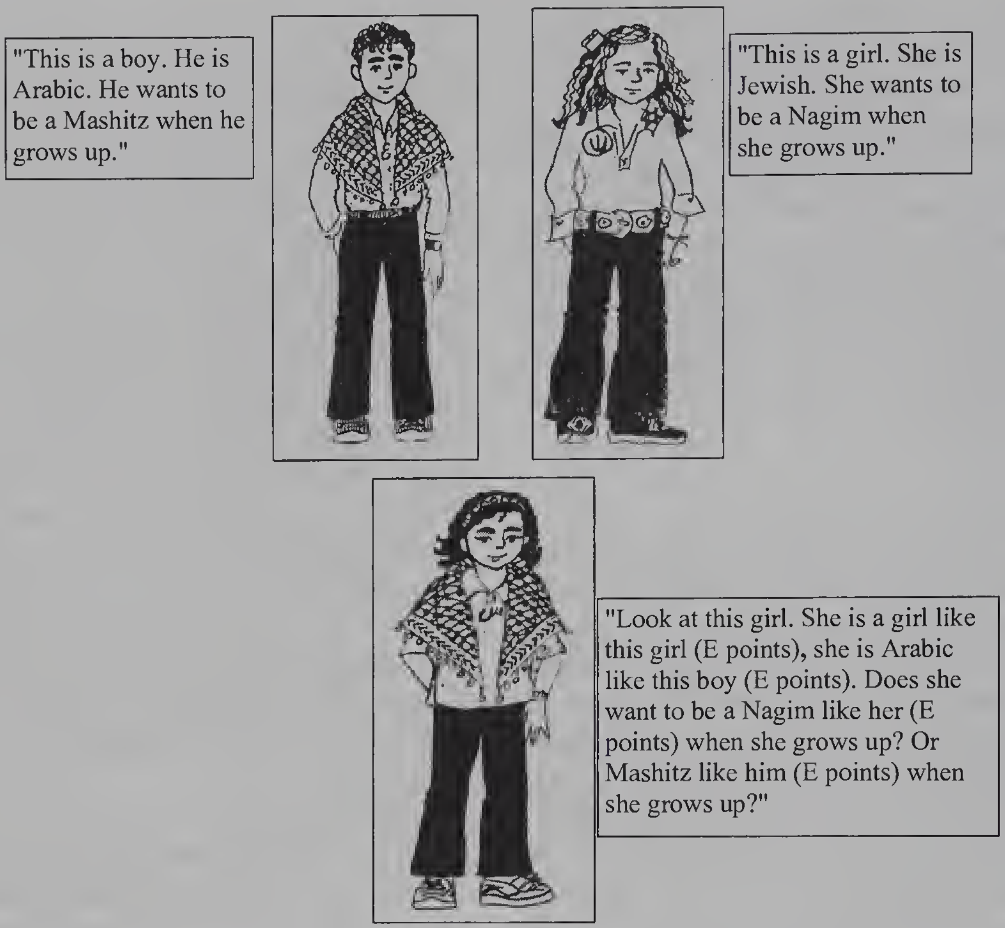

"This is a boy. He is Arabic. He wants to be a Mashitz when he grows up."
"This is a girl. She is Jewish. She wants to be a Nagim when she grows up."
"Look at this girl. She is a girl like this girl (E points), she is Arabic like this boy (E points). Does she want to be a Nagim like her (E points) when she grows up? Or Mashitz like him (E points) when she grows up?"

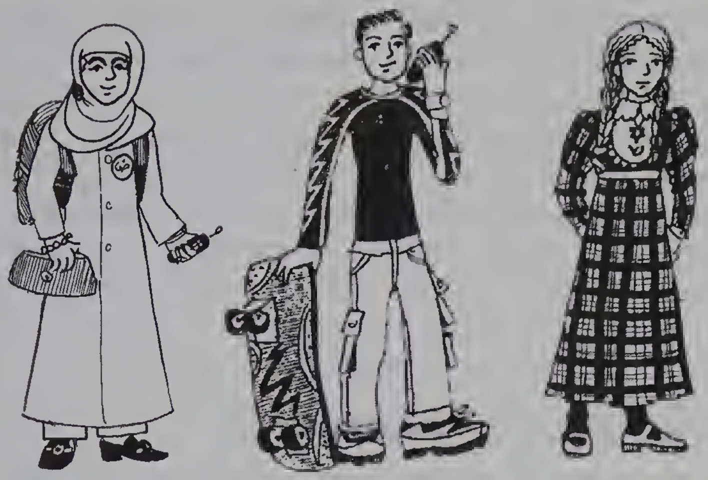
Figure 1. Example of an ethnicity versus gender triad in the relative power part of Study 1.
Note. The character on the bottom is the target character. The character on the top left is the ethnicity match. The character on the top right is the gender match.

irrelevant, either by equating them (e.g., height) or by randomizing them across all characters (e.g., hair color). Moreover, in all triads in which gender was not a target social category, the gender of all characters was the same.

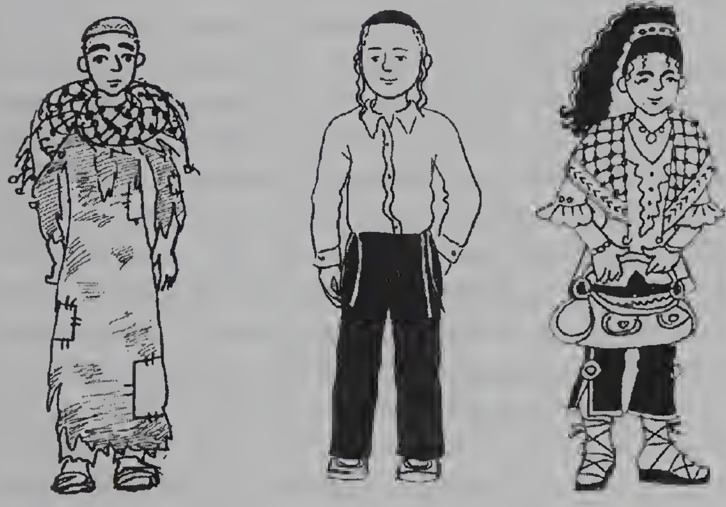
Each triad (e.g., ethnicity/gender) was presented to each participant on two separate trials. Each time, the target character was of a different value (e.g., for one child, on one of the two trials with this triad, the target character was an Arabic/girl, and on the other trial, the target character was a Jewish/boy). All possible target characters composed of the crossing of all social categories were presented an equal number of times across children (e.g., another child would get the target characters Arabic/boy and Jewish/girl, for the ethnicity/gender triad). The pairings of characters were randomly assigned to participants, and they appeared the same number of times across children.

The absolute power part of the study resembled the relative power part, except that each character was described by a social category and a personality trait rather than by two social categories. One of the test characters was presented to children as being of the same social category as the target but as having a different personality trait (*social category match*), while the other test character was described as having the same personality trait as the target character but as belonging to a different social category (*personality trait match*). Given that personality traits cannot be depicted visually in a conventional and exclusive manner, they were presented only via verbal labels. The visual cues indicating the social categories were the same as in the relative power part. All possible pairings of social category and personality trait were used for each child. In other words, for each child, each social category (e.g., social status) was paired with two different personality traits, such that each social category value (e.g., rich and poor) was paired with a different personality trait. The pairings were randomly assigned to participants, and they appeared the same number of times across children.

For both parts, all the properties attributed to the characters were novel, so that children could not base their answers on stigmas tying particular properties to categories. The properties had to do with preferences, patterns of behaviors, aspirations,



Rich Religious (for Arab sector) **Rich Boy** **Religious Jew**



Poor Religious (for Arab sector) **Religious Boy** (for Jewish sector) **Rich Arab**

Figure 2. Sample of pictures used in the studies.

and physiological properties, and were framed within sentences that allowed children to understand their general meaning. There were 20 different novel properties (12 for the relative power part of the study and 8 for the absolute power part), so that on each trial children were asked to draw inferences about a different property. The novel words were Hebrew-sounding words for the Jewish children (see the list below) and Arabic-sounding words for the Arab children. The full list of properties was: eats Dabo/Fifo for lunch, plays Zigo/Zabo at the school yard, pretends that s/he is a Talan/Falkan, listens a lot to Rendit/Mishnit music, in her/his school s/he is best in Salmon/Takton, wants to be a Mashitz/Nagim when s/he grows up, believes that in the moon there is Tugit/Pushrit, thinks it is important to be Manil/Gakish, thinks that Vitot/Bidot are the nicest animals on earth, believes that Dushon/Pitzug is the most beautiful place on earth, enjoys to read books about Tushib/Masrib, loves to watch TV programs about Dulon/Sulon, feels happy when s/he is in Nunga/Smuda, is afraid of Namtagim/Natgarim, gets angry when someone says s/he is Taluf/Mitlon, has blood type Papuch/Tirel, has Lidon/Kiron in her/his stomach, feels bad when s/her eats Hamgal/Lamdash, the smell of Miker/Lakanf makes her/him sneeze, has too little Vitamin G/R in her/his body.

Procedure

A female experimenter tested participants individually in a quiet area of their kindergarten or school. All Jewish children were tested in Hebrew by Jewish research assistants, and all Arab children were tested in Arabic by Muslim Arab research assistants from the same area that children were from. The relative and absolute power parts were conducted one after the other, with a brief intermission between them when testing kindergarteners. The procedure lasted between 15 and 20 min.

Relative power part. The procedure started with the experimenter placing one of the test pictures of the first trial in front and to the right of the child, telling the child the character's two social categories, and then informing the child about a novel property this test character had. In the same manner, the experimenter placed the second test character to the left of the child, assigning to this character his or her corresponding social categories and a different value for the novel property. The experimenter then placed the target character in between the two test characters, and slightly closer to the child. The experimenter told the child the target character's two social categories, explicitly noting the similarities to each of the test characters. Finally, the experimenter asked the child to infer whether the target character had the same property as the first test character or as the second test character. Figure 1 depicts a schematic example of the procedure. Once children responded by pointing to one of the test characters, the experimenter removed all the pictures, and brought out the next set of pictures. The experimenter avoided giving any feedback to the child that could be interpreted in terms of accuracy in their performance.

There were 12 trials in this part: 2 trials with each triad type. The order of presentation of the triads for each participant was random and was counterbalanced across participants within a sample (i.e., there were 16 different presentation orders). The right–left placement of the test characters, and consequently the order of presentation of the test characters, was also counterbalanced within and between subjects. For instance, for the ethnicity/gender triad, on one trial the gender match was presented first and was placed to the right of the child, and on the other trial, the gender match was presented second and placed to the left of the child. Finally, the order in which the social category labels were applied to each character was also counterbalanced within and between subjects. For instance, on one trial of the ethnicity/gender triad, the experimenter first noted the ethnic membership of the character and then the gender, and on the other trial she did the opposite (e.g., ''He is a boy, he is Jewish'' vs. ''He is Jewish, he is a boy''). It is important to note that in the sentences we used in Hebrew and Arabic, all the social category labels were morphologically and syntactically ambiguous as to whether they were adjectives or nouns—a fact that is representative of Hebrew and Arabic in general. Here, we translated the labels into the form that seemed most appropriate for the English construction.

Participants' responses in the first part were scored according to the number of trials in which children drew inferences based on a particular social category. Given that each triad of social categories appeared twice for each participant, scores for each triad could vary from 0 to 2 (e.g., on ethnicity/gender triads, children could select the ethnicity match on 0, 1, or both of the trials). These Triad Relative Inference Scores were used in the analyses of children's preferences for specific social categories. In order to evaluate potential differences across sectors and age groups, we summed all such scores for any given social category, creating Total Relative Inference Scores. Given that each social category appeared on three triads, with each triad appearing twice, the Total Relative Inference Score for each category could range from 0 (*no inferences based on this category*) to 6 (*inferences based on this category in all trials involving this category*).

Absolute power part. The procedure in the absolute power part was similar to the procedure in the relative power part, except that here each character was described by a social category and a personality trait (e.g., ''nice Arab,'' ''shy Jew''), and there were only eight trials. The same counterbalancing precautions described above with regard to the relative power part were also obeyed in this part. Children's responses in this part were scored according to the number of trials in which children drew inferences based on a social category or a personality trait. We refer to these scores as the Absolute Inference Scores. Given that each social category appeared on two trials in this part, the Absolute Inference Score for each category could range from 0 (*no inferences based on this category*) to 2 (*inferences based on this category in both trials involving this category*).

Results

Comparisons Across Social Categories

In order to compare the inductive power of social categories, we compared the frequency of children's selections of the different social categories to chance, conducting separate analyses for each of the two parts of the study.

On the relative power part, to avoid violating an assumption of independence of measurements, and given that the Triad Relative Inference Scores ranged from 0 to 2, we conducted nonparametric tests assessing the distribution of children on each triad type. For instance, we counted the number of children, out of the total 144, who on the two ethnicity/gender triads picked the ethnicity match on both, only one, or none of the trials. The expected (chance) distribution was 25% (36 children), 50% (72 children), and 25% (36 children), respectively. These analyses revealed that the only triad types in which the distribution significantly deviated from chance were ones involving ethnicity. In particular, on the ethnicity/gender triads, significantly more children than were expected by chance selected ethnicity on both trials (distribution for both, one, and none, was 57, 62, and 25, respectively), $\chi^2(2, N = 144) = 17.00$, $p < .001$. A similar pattern was found on ethnicity/religiosity triads (distribution for both, one, and none, was 48, 75, and 21, respectively), $\chi^2(2, N = 144) = 10.38$, $p < .01$. On ethnicity/social status triads the pattern was in a similar direction, but not significantly different from chance (distribution for both, one, and none, was 46, 68, and 30, respectively), $\chi^2(2, N = 144) = 4.00$, $p = .13$. On all other triad types the distributions did not differ from chance, $ps > .2$.

Similar tests were conducted on the absolute power part, in which each social category appeared twice, each time paired against a different personality trait. The analyses showed that only ethnicity was selected more often than expected by chance (distribution for both, one, and none, was 37, 84, and 23, respectively), $\chi^2(2, N = 144) = 6.72$, $p < .05$.

Altogether, these findings reveal that across all sectors and ages, ethnicity trumped alternative social categories and personality traits as the strongest basis for induction.

Comparisons Across Sectors and Age Groups

The second main goal of Study 1 was to compare the inductive power of various social categories across Israeli children from different ages and sectors. In order to maximize the representativeness of this assessment, we calculated a Total Inference Score for each category by adding each category's Total Relative Inference Score (0–6) and Absolute Inference Score (0–2). These scores could thus range from 0 (*no inferences based on the category*) to 8 (*all inferences based on the category*).

We entered the Total Inference Scores for each social category into a multivariate analysis of variance (MANOVA), with age (kindergarten, second grade, and sixth grade) and sector (Arabs, secular Jews, and religious Jews) as between-subjects factors (see Figure 3 for means). The overall MANOVA revealed a significant effect of sector, $F(8, 266) = 2.19$, $p < .05$, $\eta^2 = .06$, but neither the effect of age nor the interaction between sector and age were significant, *p*s > .09. Looking at each social category separately, we found that sector had a significant effect on ethnicity, $F(2, 135) = 4.21$, $p < .05$, $\eta^2 = .06$. Tukey's honestly significant difference (HSD) post hoc tests revealed that religious Jews made ethnicity-based inferences more than Arabs, $p < .05$. Sector also had a significant effect on social status, $F(2, 135) = 3.39$, $p < .05$, $\eta^2 = .05$, with Arab children drawing the most inferences based on social status and secular Jewish children the least (Tukey's HSD, $p < .05$). We also found that age had a significant effect on gender, $F(2, 135) = 3.97$, $p < .05$, $\eta^2 = .05$, with kindergarteners drawing the most inferences based on gender, and sixth graders the least (Tukey's HSD, $p < .05$).

In order to further evaluate children's inferential tendencies, we compared the Total Inference Score of each social category, for each age group and sector, against chance (chance = 4). As can be seen in Figure 3, these analyses revealed distinct sectorial and developmental patterns of responses. Among Arab children, the only significant finding was that sixth graders drew fewer inferences based on gender than would be expected by chance. Among secular Jews, none of the Total Inference Scores, at any age, were significantly different from chance. Finally, among religious Jews, the pattern was quite different, as children at all ages made ethnicity-based inferences—and only ethnicity-based inferences—significantly more than would be expected by chance.

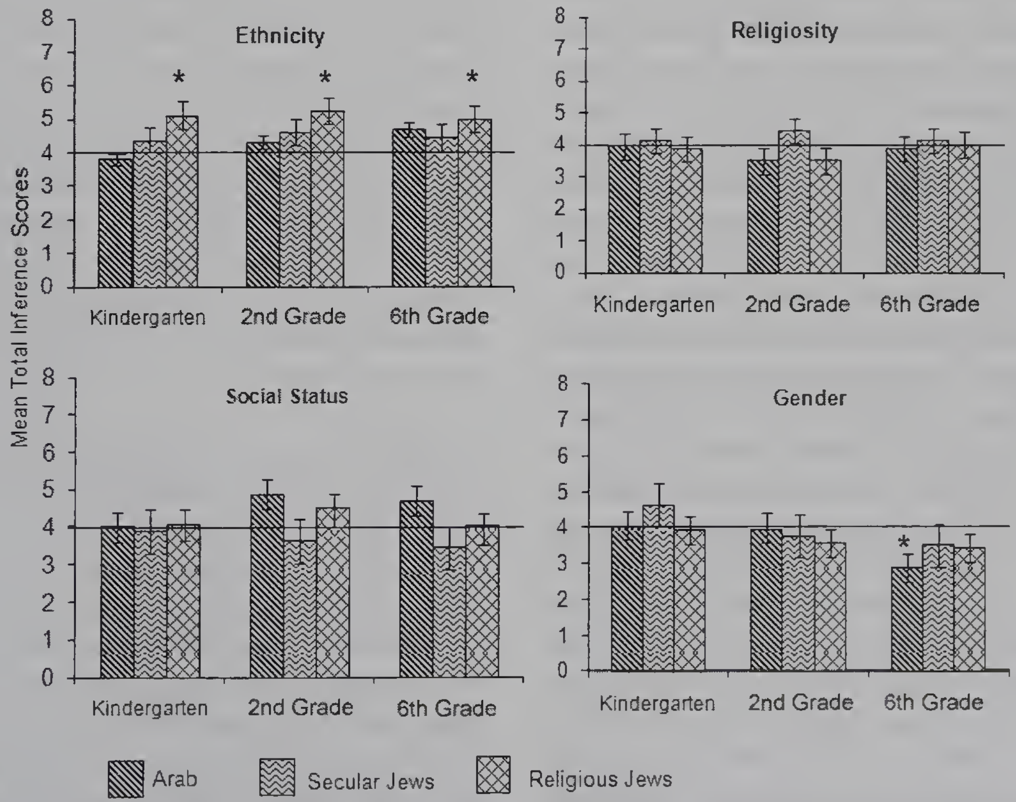


Figure 3. Mean Total Inference Scores for each social category, age group, and sector, in Study 1.
Note. The horizontal bold line indicates chance value (4). Asterisks denote difference from chance, on *t* tests (15), $p < .05$.

Analyses of Group Membership

Given the structure of the stimulus set, the study also allowed us to explore an additional set of issues having to do with group membership. In particular, we analyzed whether there were differences in the pattern of participants' inferences, based on whether the target of the inferences belonged to the same group as the participant. For this purpose, we conducted two sets of analyses, one looking at the ethnic membership of the participants (Jew or Arab), and another looking at the gender membership (boy or girl).

For the analyses on ethnicity, we created two new variables: (a) the number of trials in which children drew inferences based on ethnicity when the target character was a Jew and (b) the number of trials in which children drew inferences based on ethnicity when the target character was an Arab. Given the counterbalancing precautions noted before, for each participant, half of all eight trials involving ethnicity across the two parts of the study had a Jewish character as target, and half had an Arab character as target. Thus, these scores ranged from 0 to 4. For the analyses, we divided the participants into two sectors according to their ethnicity: Arabs and Jews (including both secular and religious). (Separate analyses on religious and secular participants generated the same results as the ones reported below.)

The above scores were entered into a repeated measures analysis of variance (ANOVA) in which target type (Jew, Arab) constituted a within-subjects factor, and ethnic sector (Jews, Arabs) a between-subjects factor. This analysis revealed only a significant interaction between target type and ethnic sector, $F(2, 138) = 5.27$, $p < .05$, $\eta^2 = .04$. To uncover the source of this interaction, we conducted separate paired t tests within each ethnic sector, comparing between the inference scores for each target type. These analyses revealed that while among Jewish participants, there was no difference between their tendency to draw inferences based on ethnicity when the target character was a Jew ($M = 2.49$, $SD = 1.19$) versus when the target character was an Arab ($M = 2.32$, $SD = 1.14$), $p > .3$, Arab participants were less likely to draw inferences based on ethnicity when the target character was a Jew ($M = 1.89$, $SD = 0.95$) than when the target character was an Arab ($M = 2.35$, $SD = 0.93$), $t(47) = 2.44$, $p < .05$. As can be seen by the means, the exceptional pattern of inferences stemmed from the Arab participants' responses to Jewish target characters.

Analogous analyses were conducted with regard to gender. For these analyses, we created two new scores: (a) the number of trials in which children drew inferences based on gender when the target character was a boy (out of 4) and (b) the number of trials in which children drew inferences based on gender when the target character was a girl (out of 4).

These scores were entered into a repeated measures ANOVA in which target type (Boy, Girl) constituted a within-subjects factor, and participants' gender (boys, girls) a between-subjects factor. This analysis revealed a significant effect of target type, $F(1, 138) = 8.88$, $p < .005$, $\eta^2 = .06$, such that children were overall more likely to draw inferences based on gender when the target character was a girl ($M = 2.11$, $SD = 1.19$), than when the target character was a boy ($M = 1.67$, $SD = 1.15$). This main effect was subsumed by a close to significant interaction between target type and participants' gender, $F(1, 138) = 3.14$, $p = .079$, $\eta^2 = .02$. Given the exploratory nature of this analysis, and the analogy with the results on ethnicity, we followed up this analysis by again conducting separate paired t tests within each participants' gender group, comparing between the inference scores for each target type. These analyses revealed that while among male participants, there was no difference between their tendency to draw inferences based on gender when the target character was a boy ($M = 1.89$, $SD = 1.15$) versus when the target character was a girl ($M = 2.00$, $SD = 1.25$), $p > .6$, female participants were less likely to draw inferences based on gender when the target character was a boy ($M = 1.53$, $SD = 1.9135$) than when the target character was a girl ($M = 2.18$, $SD = 1.16$), $t(47) = 3.93$, $p < .001$. As can be seen by the means, the exceptional pattern of inferences stemmed from the female participants' responses to male target characters.

Comparison of Social Categories Against Personality Traits

A minor goal of Study 1 was to assess the development of children's overall social category-based induction. Earlier studies conducted among secular Jews in Israel had found that while kindergarteners were more likely to draw inferences based on social category membership than based on personality traits, adults showed the reverse pattern (Diesendruck & haLevi, 2006). The absolute power part of Study 1 allowed us to assess this developmental trend in more detail. For this purpose, we computed a Social Categories Inference Score by simply adding the Absolute Inference Scores across the four social categories. This score could range from 0 (*all inferences based on personality traits*) to 8 (*all inferences based on social categories*). We then conducted an ANOVA on these Social Categories Inference Scores, with age and sector as between-subjects factors. The only close to significant effect was of age, $F(2, 135) = 2.81$, $p = .06$, $\eta^2 = .04$. Given our interest in the development of this tendency, we compared the Social Categories Inference Score for each age group separately to chance (chance = 4). These analyses revealed that both kindergarteners ($M = 4.67$, $SD = 1.50$), $t(47) = 3.07$, $p < .005$, and second graders ($M = 4.67$, $SD = 1.62$), $t(47) = 2.86$, $p < .01$, drew inferences based on social categories significantly more than would be expected by chance. In turn, sixth graders did so at a level no different than chance ($M = 4.02$, $SD = 1.56$), $p = .9$. In other words, it seems that by sixth grade, children are midway toward the adult pattern of reliance on personality traits.

Discussion

The results of Study 1 reveal that, overall, ethnicity was the most powerful source of induction. Importantly, it appears that ethnicity was especially salient for religious Jewish children, and it became a powerful source of induction earlier for them than for children from the other two sectors. In other words, while there are some indications that by sixth grade, children from other sectors (e.g., Arabs) start construing ethnicity as the most powerful source of induction among a variety of human categories, religious Jewish children manifest this bias already by kindergarten.

The above conclusions raise questions about why is it that ethnicity is such a powerful source of induction for children in Israel, and why is it especially important for religious Jewish children. Some of the plausible responses to these questions are beyond the scope of the present studies, and will be raised in the General Discussion. Nonetheless, a direct question has to do with whether ethnicity's inductive power derived from the salience of its visual markers. That is, was the presence of the unique attire and visual symbols associated with the Jewish and Arab characters in our materials, a more visually compelling cue to children than the other social category markers, making that dimension of similarity stand out above all others? Were these cues especially salient for religious Jewish children? Or did the inductive power of ethnic categories derive from their linguistic markers (i.e., labels), and thus exerted their effect independently of children's capacity to visually recognize the categories? Study 2 addressed these alternatives.

Study 2

In order to assess the relative contributions of visual versus linguistic markers to the effects found in Study 1, Study 2 adopted the same procedure as in the relative power part of Study 1, except that the experimenter did not provide children any social category labels. Thus, in Study 2, the only information available to children about the social category membership of the characters displayed was their physical markers. The rationale was that if the inductive power of ethnicity—or any other social category—derives from bottom-up perceptual processes, then sheer exposure to stereotypical visual markers of ethnic categories should lead children to draw inferences based on them. Moreover, we should find that religious Jewish children will show the same higher sensitivity to ethnicity as found in Study 1. Alternatively, if the inductive power of ethnicity—as revealed in Study 1—is attributable primarily to the presence of cultural markers such as labels, then children in Study 2—including religious Jewish children—should not draw inferences based on ethnicity when only physical information about category membership is available. A comparison between children's responses in the two studies would allow us to estimate directly the impact of labeling on children's inferences.

In a sense, this comparison complements one reported in Diesendruck and haLevi (2006). In that study, children were tested in one of two conditions: a category labels + visual markers condition—similar to the present Study 1—and a category labels, no visual markers condition. That study thus assessed the relative contribution of *visual markers* to the inductive potential of various social categories. In fact, Diesendruck and haLevi found that visual markers did not contribute significantly to the inductive power of social categories. The present Study 2 assessed the complementary effect by adding the missing cell, that is, a visual markers-only condition. It thus allowed us to evaluate how much *labels* contribute to the inductive potential of social categories.

In Study 2, we did not replicate the absolute power part of Study 1, because that part contrasted social categories with personality traits. While one can visually hint at the personality trait of a person by, for instance, illustrating his or her behavior, we believe these visual cues differ from those associated with most of the social categories examined here insofar as they are not conventional and stereotypical physical markers that exclusively identify the category membership of an individual.

In addition, in order to estimate potential differences in the salience or recognizability of the social categories, we conducted a ''category recognition'' task at the end of the triad induction task. Children's accuracy would allow us to evaluate to what extent differences in the inductive power of social categories and across sectors could be attributed to the ease of category recognition.

Method

Participants

A total of 144 children participated in this study (see Table 1 for details). The samples and recruitment methods were identical to those in Study 1. None of the children had participated in Study 1.

Design

Study 2 consisted of two parts: an induction part and a category recognition part. The induction part was nearly identical to the relative power part of Study 1, except that the experimenter did not provide category labels for any of the characters. The same 12 trials presented in Study 1 were presented in this part of Study 2. In the category recognition part, children were shown a series of six triads that had appeared in the induction part and, on separate trials, were asked to describe and identify the characters corresponding to the different social category labels mentioned by the experimenter.

Materials

The materials were the same as those of the relative power part of Study 1.

Procedure

The procedure was identical to the one in Study 1. The category recognition part followed the induction part after a short intermission between them.

Induction part. The procedure in the induction part was identical to the procedure in the relative power part of Study 1, except that the experimenter did not mention the social categories of each character. Children's responses were scored in the same way as in Study 1: namely, according to the number of trials (0–2) within each triad type in which they selected a particular social category (Triad Relative Inference Scores) and according to the total number of trials (0–6) across triad types in which children drew inferences based on a particular social category (Total Relative Inference Scores).

Category recognition part. Category recognition was assessed in both an unprompted and a prompted manner. For the two unprompted trials, the experimenter showed children the six pictures from the first two trials of characters that children had seen in the induction part. Children were simply asked to describe each character in their own words, and the experimenter wrote down these descriptions. After these two unprompted trials, the experimenter moved on to the prompted trials, which tested children's capacity to recognize the characters as exemplars of the target categories. For this purpose, the experimenter laid down the three pictures of the third trial each child had seen in the induction part and asked the child to identify the referent characters of the different social categories depicted in those pictures. For example, regarding the triad displayed in Figure 1, the experimenter asked: ''Is there an Arab among these characters? Who is an Arab?'' The experimenter wrote which characters the child recognized as Arabs and then asked about the next categories depicted (e.g., ''Is there a Jew among these pictures? Is there a girl? A boy?''). The same procedure was repeated for the next three triads in the child's set, thus resulting in the presentation of a total of 12 pictures.

Children's responses in the unprompted trials were scored according to the number of times in which children explicitly mentioned the target category value when describing the character (e.g., said ''Jew'' of a Jewish character). We refer to these as the Unprompted Category-Recognition Scores. Children's responses in the prompted trials were scored according to the number of characters identified correctly for each social category. Given that in this part, each social category appeared on six pictures, children's Prompted Category-Recognition Scores for each social category varied from 0 (*no category exemplar correctly identified*) to 6 (*all category exemplars correctly identified*).

Results

Induction Part

As in Study 1, we first compared the inductive power of the various social categories and then analyzed the effects of sector and age group on children's choices.

Comparisons across social categories. As in Study 1, in order to evaluate the inductive power of each social category, we analyzed the distribution of children according to their Triad Relative Inference Scores on each triad type, comparing it against the expected chance distribution (36, 72, and 36, respectively, for selecting a particular dimension on both, one, or none of the trials of a given type). These analyses revealed a pattern of results different from the one found in Study 1. In particular, on ethnicity/gender triads, again significantly more children than were expected by chance selected ethnicity on both trials (distribution for both, one, and none, was 53, 58, and 33, respectively), $\chi^2(2, N = 144) = 11.00$, $p < .005$. However, on ethnicity/social status triads, the pattern was the reverse. Namely, significantly more children selected ethnicity on *neither one* of the two trials (i.e., selected social status on both) than what was expected by chance (distribution for both, one, and none, was 22, 70, and 52, respectively), $\chi^2(2, N = 144) = 12.61$, $p < .005$. Social status was also frequently selected on both trials in social status/gender triads (distribution for both social status, one, and none, was 65, 46, and 33, respectively), $\chi^2(2, N = 144) = 33.00$, $p < .001$. On all other triad types the distributions did not differ from chance, *ps* > .2. In sum, when no labels were offered, children did not consistently privilege ethnicity over all other social categories and, in fact, seemed to privilege another social category—social status—instead.

Comparisons across sectors and age groups. In order to evaluate the effects of sector and age group on children's choices, we entered the Total Relative Inference Score for each social category into a MANOVA, with age (kindergarten, second grade, and sixth grade) and sector (Arabs, secular Jews, and religious Jews) as between-subjects factors (see Figure 4 for means). The overall MANOVA revealed only a significant effect of age, $F(8, 266) = 4.33$, $p < .001$, $\eta^2 = .12$. Neither the effect of sector nor the interaction between sector and age was significant, *ps* > .4. Looking at each social category separately, we found that age had a significant effect on ethnicity, $F(2, 135) = 4.56$, $p < .05$, $\eta^2 = .06$, with children making more ethnicity-based inferences as they grew older (difference between sixth and second graders, $p < .05$, Tukey's HSD). Age also had a significant effect on social status, $F(2, 135) = 3.33$, $p < .05$, $\eta^2 = .05$, again with children making more social-status-based inferences as they matured (difference between sixth graders and kindergarteners, $p < .05$, Tukey's HSD). Finally, age also had an effect on gender, $F(2, 135) = 12.36$, $p < .001$, $\eta^2 = .16$, with sixth graders being less likely to make gender-based inferences than both second graders and kindergarteners (Tukey's HSD, *ps* < .001). Neither sector nor the interaction between sector and age had significant effects on any of the individual social categories' Total Relative Inference Scores.

In order to further evaluate children's inferential tendencies, we compared the Total Relative Inference Score of each social category, for each age group and sector, against chance (chance = 3). As can be seen in Figure 4, these analyses revealed sectorial and developmental patterns of responses that were quite different from the ones found in Study 1 (compare asterisks in Figures 3 and 4). Among Arab children, in kindergarten, none of the scores were significantly different from chance; in second grade, religiosity was selected significantly less often than expected by chance, and in sixth grade, ethnicity was selected significantly more than expected by chance, and gender significantly less than expected by chance. Among secular Jews, the only score that differed significantly from chance was for gender in sixth grade, with children drawing gender-based inferences less often than expected by chance. Finally, among religious Jewish children, only by sixth grade, children were more likely to draw inferences based on social status than would be expected by chance and were less likely to draw inferences based on gender than would be expected by chance.

Comparisons Between Studies 1 and 2

In order to assess the effect of labels on children's responses, we conducted a MANOVA on the Total Relative Inference Scores of each social category in Study 2 (no labels) and Study 1 (with labels). Recall that the relative power part was identical in the two studies, with the only difference being the presence of category labels in Study 1. In this analysis, we entered study, age, and sector as between-subjects variables. Here we report only on the effects involving Study.

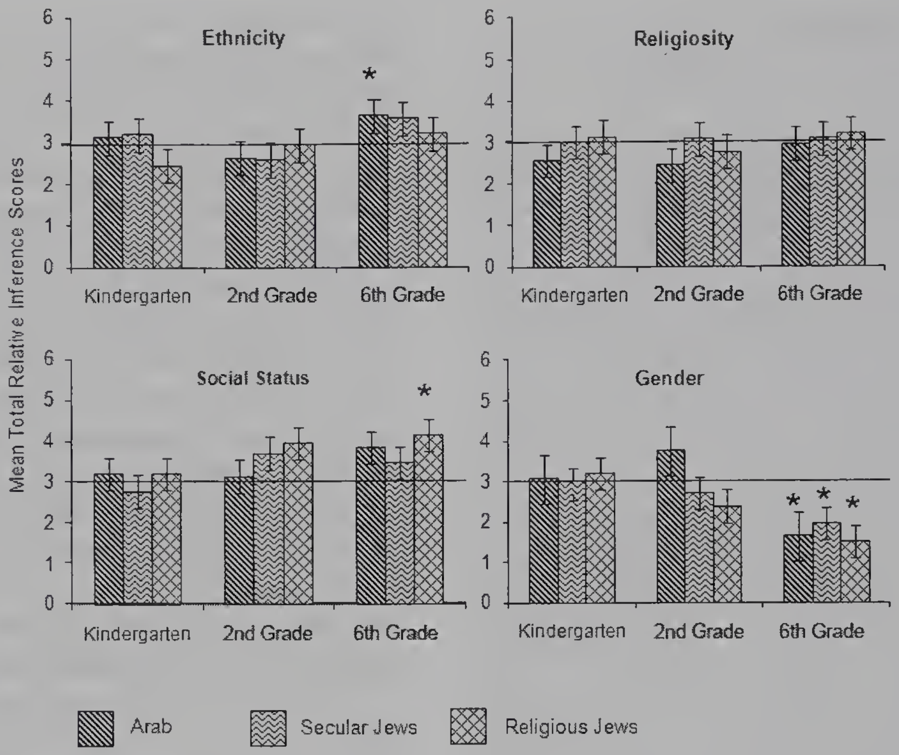


Figure 4. Mean Total Relative Inference Scores for each social category, age group, and sector, in Study 2.
Note. The horizontal bold line indicates chance value (4). Asterisks denote difference from chance, on *t* tests (15), $p < .05$.

The overall MANOVA revealed a main effect of study, $F(4, 267) = 3.85$, $p < .01$, $\eta^2 = .06$ (see Figure 5). The separate ANOVAs showed that this effect was significant for ethnicity, $F(1, 270) = 10.03$, $p < .005$, $\eta^2 = .04$, with children *more* likely to draw ethnicity-based inferences in Study 1 ($M = 3.52$, $SD = 1.39$) than in Study 2 ($M = 3.03$, $SD = 1.28$). The effect of study was also significant with regard to social status, $F(1, 270) = 8.11$, $p < .01$, $\eta^2 = .03$, with children *less* likely to draw social-status-based inferences in Study 1 ($M = 2.99$, $SD = 1.43$) than in Study 2 ($M = 3.47$, $SD = 1.49$). Finally, the interaction between age and study was significant with regard to gender, $F(2, 270) = 3.40$, $p < .05$, $\eta^2 = .03$, and the interaction between sector and study was significant with respect to ethnicity, $F(2, 270) = 3.18$, $p < .05$, $\eta^2 = .02$.

In order to identify the source of the theoretically central Sector × Study interaction on ethnicity, we conducted separate ANOVAs on each of the three sectors. In these analyses, the Total Relative Inference Scores for ethnicity was used as the dependent measure, and study as a between-subjects factor. These ANOVAs revealed that while study did not have a significant effect on either Arab or secular Jewish children's tendency to draw inferences based on ethnicity (*ps* > .3), study did have a significant effect on religious Jewish children's tendency to do so, $F(1, 94) = 14.19$, $p < .001$, $\eta^2 = .13$. In particular, these children were more likely to rely on ethnicity when it was labeled (Study 1, $M = 3.90$, $SD = 1.48$) than when it was not (Study 2, $M = 2.85$, $SD = 1.22$).

Category Recognition Part

The main goals of the category recognition part of Study 2 were to estimate potential differences in the visual salience of the different social categories, and to evaluate whether there are differences in this respect between ages and sectors.

Preliminary analyses of the unprompted trials revealed no significant differences between the values within each category (e.g., no differences between children's spontaneous recognition of Jews vs. Arabs). We thus analyzed the data using as the dependent measures children's Unprompted Category-Recognition Scores for each of the three social categories. Given that gender is pervasively marked in both Hebrew and Arabic, it would have been pragmatically odd for children to mention a character's gender category. For this reason, we excluded the category gender from the analyses.

We entered these scores into a MANOVA with age and sector as between-subjects variables. This analysis revealed only a significant effect of age, $F(6, 244) = 11.89$, $p < .001$, $\eta^2 = .23$. Neither the effect of sector nor the interaction between age and sector was significant (*ps* > .2). The individual ANOVAs on each category revealed that age had a

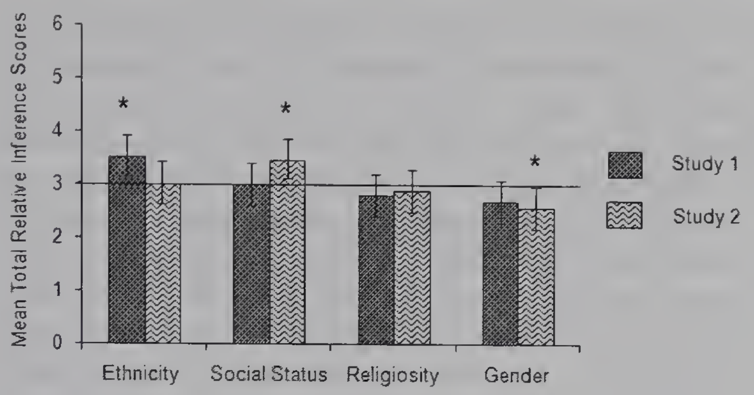


Figure 5. Comparison of mean Total Relative Inference Scores between studies, for each social category.
Note. The horizontal bold line indicates chance value (3). Asterisks denote difference from chance.

significant effect on all categories ($ps < .001$), with sixth graders more likely to mention the category labels than children from the other two age groups (Tukey's HSD, $ps < .001$). Additionally, we conducted paired t tests comparing the frequency with which children mentioned the category labels across the different categories. These tests revealed that the proportion of trials in which children mentioned the social status label of the characters ($M = 0.27$, $SD = 0.38$) was significantly higher than the proportion of trials in which children mentioned either a religiosity label ($M = 0.16$, $SD = 0.26$), $t(134) = 3.76$, $p < .001$, or an ethnicity label ($M = 0.18$, $SD = 0.34$), $t(132) = 2.90$, $p < .005$.

For the analyses of the prompted trials, we compared the Prompted Category-Recognition Score for each social category (range = 0–6), and in each age group and sector, against chance (chance = 3 pictures recognized correctly). The Gender Recognition Scores were close to ceiling (above 5) in all ages and sectors, and were therefore omitted from the analyses. As depicted in Figure 6, these analyses showed that in kindergarten, Arab children identified ethnicity at a rate significantly lower than expected by chance, and religiosity and social status at chance. Secular Jewish children identified religiosity at a rate lower than expected by chance, and ethnicity and social status at chance. Finally, religious Jewish children identified all categories at chance. By second grade, and even more so at sixth grade, all children identified all categories at a rate better than expected by chance.

Last, to directly address potential associations between children's capacity to recognize members of a social category, and children's tendency to draw inferences based on a social category, we assessed the correlations between children's Prompted Category-Recognition Score on any given social category, and their Total Relative Inference Score on these same categories. None of these correlations was significant ($ps > .16$).

Discussion

One of the main goals of Study 2 was to assess whether the presence of visual markers of social category membership was sufficient to generate the

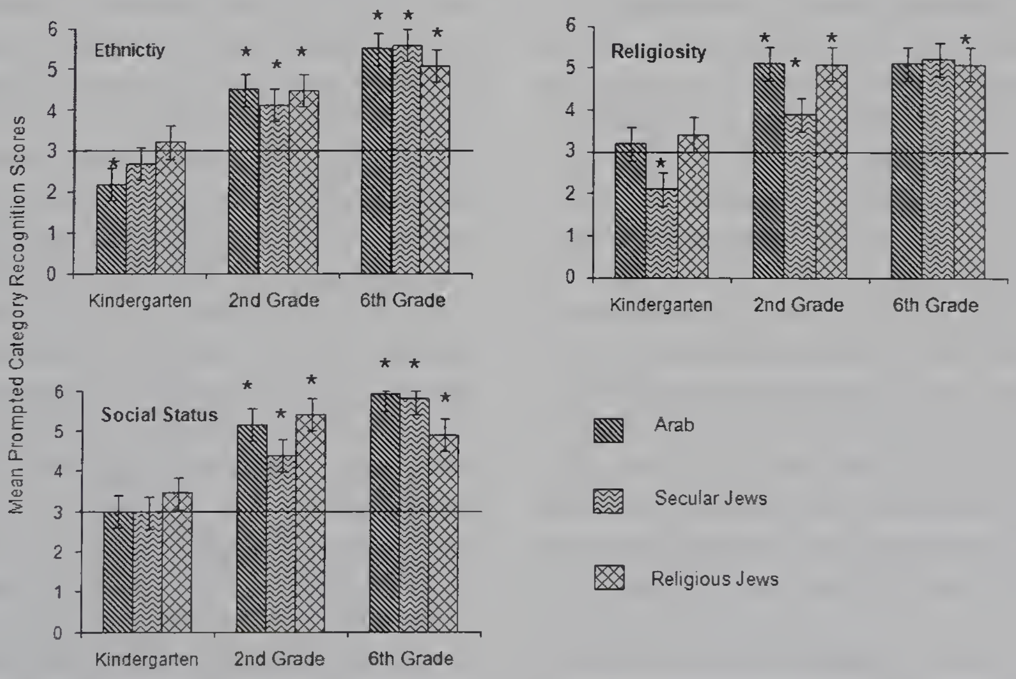


Figure 6. Mean Prompted Category Recognition scores for each social category, age group, and sector, in Study 2.
Note. The horizontal bold line indicates chance value (4). Asterisks denote difference from chance, on t tests (15).
$p < .05$.

pattern of induction observed in Study 1. The simple answer is "no." The most consistent findings from Study 2 were that with development, children's tendency to draw inferences based on social status and ethnicity increased, and their tendency to draw inferences based on gender decreased. Crucially for the present purposes, however, Study 2 revealed neither a consistent effect of sector on children's inferences nor a distinctive sensitivity to ethnic categories. In particular, there was no evidence that religious Jewish children were more likely to draw ethnicity-based inferences than children from the other sectors.

The comparison between the two studies revealed a number of important patterns. First, they confirmed that with development, children's reliance on gender as a source for induction declines, but their reliance on ethnicity rises. These results on gender are consistent with Taylor's (1996) finding of a decrease, from kindergarten to fifth grade, in children's beliefs about the innateness of gender-typical behavior. Second, the analyses showed that labeling has a distinctive effect on children's ethnicity-based inferences. In particular, labeling significantly increased the frequency of ethnicity-based inferences but had no effect on either religiosity- or gender-based inferences. There was an effect of labeling on social-status-based inferences but in a negative direction; that is, labeling decreased social status-based inferences. Given that the task required children to choose between social categories, it is possible that the decrease in social-status-based inferences in Study 1 was partly on account of the increase in ethnicity-based inferences. It is nonetheless revealing that the increase was only in terms of ethnicity. Third and finally, the analyses by sector showed that labeling had a significant effect only among religious Jewish children. In other words, labeling in itself does not seem to create an inferential bias in a vacuum. Rather, it is only within a population already disposed toward such a bias that labeling carries an effect. We will return to these issues in the General Discussion.

The findings from the category recognition part of Study 2 further confirmed the overall conclusions noted above. In particular, first, we found that there were no differences across children from the different sectors in their capacity to recognize social category members in the pictures we used as stimuli. In other words, any differences that exist between children from the different sectors in terms of the inductive power of a given social category—especially ethnicity—do not seem to reflect differences across sectors in the ability to recognize the visual cues associated with a category, or in the visual salience of a category to a particular sector. This was further confirmed at an individual child level, by the lack of significant correlations between children's recognition and induction scores. Second, we did find a consistent increase with age in children's capacity to recognize social category members, but again, this developmental effect held across all three sectors. Thus, even though in Study 1, religious Jewish children—and only them—favored ethnicity as a source of induction already by kindergarten and second grade, it does not seem that this pattern derived from the stronger perceptual salience of ethnicity to children in this sector at those ages. In fact, by second grade, children from all sectors recognized ethnic categories in the prompted trials at above chance rates, and yet this increased recognition capacity was not accompanied by an increase in the rate of ethnicity-based inferences.

General Discussion

A number of studies conducted with both children and adults have found that in different societies, different social categories seem to be essentialized (e.g., Astuti et al., 2004; Hirschfeld, 1996; Mahalingam, 2003). One of the main goals of the present studies was to assess potential differences among sectors within a given society in their tendency to essentialize the same social categories and to examine this matter from a developmental perspective. The three sectors of the Israeli society evaluated here—secular Jews, Zionist modern religious Jews, and Muslim Arabs—were selected for the presumed diversity in their perspectives regarding ethnicity. The hypotheses were that to the extent that essentialist beliefs about a given social category are influenced by the cultural input to which children are exposed, we should find differences in a measure of essentialism among children from these three sectors. Moreover, to the extent that essentialism is a bias available to children from early on, then we should find evidence for these hypothesized differences among sectors from the earliest age assessed.

The findings from Study 1 were consistent with the above sectorial and developmental hypotheses. In that study, we found that across all ages and sectors, children were most likely to draw inferences based on ethnicity. That is, overall, when characters were described as Jews or Arabs, children were more likely to focus on that categorical similarity when drawing novel inferences about these characters than on any dimension of similarity between the characters, be the dimension another social category (e.g., gender) or a personality trait (e.g., niceness). This privileged status of ethnicity seems to reflect the salience of this category in public discourse in Israel (Bar-Tal & Teichman, 2005) and is generally consistent with previous findings among Jewish secular kindergarten children (Diesendruck & haLevi, 2006). In fact, this focus on ethnicity has been found in other societies as well (see, e.g., Astuti et al., 2004; Gil-White, 2001; Kanovsky, 2007).

A further important finding from Study 1 is that there were nonetheless differences in the emphasis on ethnicity among sectors. In particular, compared to secular Jews and Muslim Arabs, we found that religious Jewish children were the most likely to draw inferences based on ethnicity. In fact, among kindergarteners and second graders, only the religious Jewish children drew inferences based on ethnicity significantly more than would be expected by chance. This differs from the findings of Diesendruck and haLevi (2006) who found distinctive attention to ethnicity also among secular Jewish kindergarteners. One possible explanation for this apparent disparity in the responses of secular children regarding ethnicity is that while in Diesendruck & haLevi, ethnicity was contrasted exclusively against personality traits, in the present study ethnicity was first contrasted against other social categories. This exposure to other social categories may have weakened the "appeal" of ethnicity to secular Jewish kindergarten children in the present study. In other words, it is possible that for secular Jewish children, ethnicity is a powerful enough category that it trumps personality traits, but it is not powerful enough to trump other social categories. Ethnicity does have this latter kind of power for religious Jewish children, intimating that this group may be the most essentialist about ethnicity among the three sectors who participated in these studies.

The above conclusion raises the question of what might give rise to these differences among the sociocultural sectors in Israeli society, or more generally, what causes differential essentialist beliefs toward a particular social category to emerge? There are a number of possibilities raised in the literature. One possibility derived from the social psychological literature (Allport, 1954) is that the degree of contact among members of different social categories may be inversely related to the degree to which the social categories are essentialized. It seems to us, however, that this factor cannot fully account for the differences we found in the present study. For instance, the religious and secular Jewish samples were recruited from the same cities, and so it does not seem to be the case that the religious Jewish children would have less contact with Arabs than the secular Jewish children. A second possibility is that essentialism may be particularly enforced within groups that wish to maintain a power differential toward a separate group. Mahalingam (2003) raised this possibility as an account of the essentialist beliefs of Brahmin Indian adults toward other castes. While this notion may indeed be applicable to the context of Israeli society, it is unlikely that already by kindergarten, religious Jewish children are aware of this ideology of power. In other words, this ideology would have to be somehow transmitted to children by means they could understand and enforce.

A third possibility that addresses this latter matter is that due to ideological values, religious Jewish adults may emphasize distinctions between ethnic groups, which children may then pick up and reproduce (see Bigler & Liben, 2007, for a discussion of such processes). One way in which adults might emphasize ethnicity is via linguistic markers, such as chronic labeling of ethnic categories, or use of generics to describe ethnic categories. Gelman and colleagues found that such cues may be related to essentialist beliefs about gender (Gelman et al., 2004), and thus a similar process may occur in the context of ethnic categories. We are currently examining this possibility. A final possibility we would like to raise is that religious beliefs per se may endorse essentialist beliefs. In fact, in a recent study among secular and religious Jewish children, Diesendruck and Haber (2009) found that children's beliefs about God as the creator of social categories was positively related to children's essentialist construal of the social categories.

Clearly, there are a number of research routes to undertake in order to uncover why a particular group within a society comes to essentialize a particular social category. The present findings not only establish the need to pursue these routes but also provide some initial guidelines for doing so. One such guideline mentioned above that carries important theoretical implications has to do with the age at which differences across sectors start emerging. In Study 1, we found that the differences between religious Jewish children and the other two groups of children in terms of privileging ethnicity as a basis for induction were evident already in the youngest age group tested, that is, kindergarteners. This finding suggests that whatever effect cultural input might have on children's essentialist beliefs, it takes place early on in development and prior to children's intensive immersion into their sector's unique educational institutions. In fact, recent studies reveal that even under conditions of arbitrary social classification, preschoolers manifest ingroup favoritism, thus corroborating the notion that processes of social categorization have significant implications very early on in development (Patterson & Bigler, 2006). This conclusion seems most consistent with the claim that rather than having to acquire essentialist beliefs per se (Fodor, 1998), children might start off with an intuitive essentialist bias, and cultural input orients this bias to specific social categories (Hirschfeld, 1996).

One important general caveat with regard to the above conclusions is that, while we have been discussing the sectorial and developmental findings in terms of ''essentialism,'' all the results come from induction tasks. Although these tasks have been used in the past as one of the components of essentialist reasoning, it is possible that not all components of essentialism pattern in the same way; that is, that they do not form a coherent construct (Gelman, Heyman, & Legare, 2007). In fact, comparing the same populations as the ones included here but on different components of essentialism, Birnbaum (2009) found that Muslim Arab children, rather than religious Jewish children, manifested the most essentialist-like response patterns. At a minimum, nonetheless, the present findings on induction invite the same kinds of questions about the origins of the status of ethnicity as the ones raised above about essentialism in general.

The second major goal of the present studies was to evaluate the contribution of labeling to the inductive potential of social categories. This was accomplished by comparing children's responses in a task where all categories were labeled (Study 1) to a task where categories were not labeled (Study 2). There have been claims about the importance of language in the generation of essentialism in general (Carey, 1995), and social essentialism in particular (Sperber, 1996). In fact, a number of studies highlight the importance of labeling to the inductive potential of categories in general (Gelman & Markman, 1987; Sloutsky, Lo, & Fisher, 2001), and social categories in particular (Diesendruck & haLevi, 2006; Gelman & Heyman, 1999).

The findings from these two studies converged onto the same conclusion regarding this issue. Namely, that labeling plays an important role in the inductive power of social categories, and that visual cues—or perceptual similarity—are not sufficient to account for the inductive potential of social categories. First, we found that different from Study 1, in Study 2 ethnicity was not the most powerful source of induction. Second, in Study 2 we found that religious Jewish children were no more likely than secular Jewish children or Muslim Arab children to rely on ethnicity for drawing inferences. Third and finally, we found that there were no systematic differences across sectors in children's capacity to recognize the social categories depicted in the stimuli, and there were no significant correlations between children's capacity to recognize a social category and children's tendency to rely on that category as a basis for induction. In fact, the finding that religious Jewish kindergarteners privileged ethnicity as a basis for induction when ethnicity was labeled, even though they were incapable of systematically recognizing Jews and Arabs in pictures, suggests that children's concept of ethnicity predates their perceptual classification. Altogether, social categories—in the present case, ethnicity—earn their inductive potential not from their visual distinctiveness, bur rather from their labels (see also Hirschfeld, 1993).

There is considerable debate regarding the mechanisms by which labeling affects induction and essentialism. Some researchers argue that young children rely on labels for induction because of the privileged status of the auditory modality (Sloutsky et al., 2001). Others claim that labels invite children to search for the commonalities among instances of a category, thus fostering the acquisition of a notion of kind (Waxman, 1999; Xu, 2002). Finally, there are suggestions about the possibility of labels causing children to believe that deep commonalities between category members exist (Carey, 1995; Gelman & Heyman, 1999). The comparison between Studies 1 and 2 provides important new insights regarding this issue. That comparison revealed that labeling significantly increased the inductive potential only of ethnic categories, and only for religious Jewish children.

The category specificity of the effect of labeling is inconsistent with any generalist account of such an effect, perhaps especially one based exclusively on low-level perceptual capacities. This specificity intimates that children are relying on more than what is made available to them in the task; that is, children seem to be relying also on their underlying concepts about the categories. As a consequence, not all *categories* are equally susceptible to this effect. The sector specificity of the effect of labeling reinforces the above conclusion, by emphasizing that not all *children* are equally susceptible to such an effect. Labeling does not seem to create essentialism out of vacuous concepts but rather from already fertile ones. The fact that this was true for the religious Jewish children across all ages also undermines the argument that labeling effects in young children are due to the early maturation of the auditory system (Sloutsky et al., 2001).

These findings about the specificity of labeling are also relevant to a recent theory expounded by Bigler and Liben (2007) about the development of intergroup concepts. According to Bigler and Liben, the psychological salience of a particular category derives primarily from four factors: perceptual discriminability, labeling, differential distribution, and implicit use. We would add that prior concepts about real-life categories may modulate the weighting of these factors. Using the effect of labeling as an example, it is only within a particular population that a certain labeled category stood out. One caveat to this conclusion is that it is possible that while we have manipulated—and thus equated—children's exposure to social category labels *in the task,* children from different sectors may nonetheless be differentially exposed to social category labels *in their daily lives.* In other words, consistent with Bigler and Liben, ethnicity may be salient to religious Jewish children because this category is more frequently labeled in their regular input. As mentioned earlier, we are currently examining this possibility.

The conclusion that seems most consistent with all these findings is that labels are particularly effective triggers of an essentialist bias that, while intuitive to all children, may be differentially latent with regard to particular social categories depending on a child's sociocultural background. In other words, the argument is that religious Jewish children in Israel may have a stronger essentialist belief about ethnicity than children from the other two sectors tested here. This belief does not derive from children's perception of the categories or from their experiences with members of the categories. Rather, as discussed earlier, aspects of these children's sociocultural input may endorse this belief. Interestingly, it is only by sixth grade that children have sufficiently fine-tuned visual schemas for the different categories, that they do not "need" labels to activate their rich concepts about the categories. Up until then, labels are one of the primary means by which this function is achieved. In other words, rather than construing ethnic categories "from the bottom up," children are doing so by first associating abstract conceptual richness with labels, and only then recognizing the referents of these labels in their social environment.

One final finding that may have important implications for the development of social essentialism came from our analyses of the effect of group membership on children's inferences. In Study 1 we found that the experimental context least conducive to ethnicity-based inferences was one in which the participant was an Arab and the target of the inference was described as a Jew. Similarly, the context least conducive to gender-based inferences was one in which the participant was a girl and the target of the inference a boy.

A line of research that may aid in the interpretation of the above pattern of results are adult social psychological studies on intergroup perceptions. Following this literature, we may describe the present pattern of findings regarding category-based inferences in terms of the perceived homogeneity of categories. The more one relies on a particular category as the basis for inference, the more one can be said to conceive of the members of the category as being similar. Thus, for instance, the asymmetry summarized above can be described as Arab children viewing "Jewishness" as more heterogeneous than "Arabness." The casting of the findings in this light makes them directly relevant to a well-documented phenomenon on adults' intergroup perceptions—the so-called outgroup homogeneity effect. The effect refers to the finding that adults tend to perceive members of an outgroup as more homogenous than members of their ingroup (see Ostrom & Sedikides, 1992, for a review). At face value, if children were susceptible to this effect, then we would have expected the exact opposite pattern of findings from the one reported here; for example, Arab children should have perceived Jewishness as more *homogeneous* than Arabness. Interestingly, nonetheless, studies in the adult social psychological literature report that under certain circumstances, there may occur a reversal of the outgroup homogeneity effect, analogous to the one found in the present study. In particular, the effect has been found to reverse when the participants belong to a minority group with whom they identify highly, and whom they positively value (Doosje, Ellemers, & Spears, 1995). To the extent that Arab children (and girls) correspond to these characteristics, the present findings may be taken as evidence for the early manifestation of this adult social psychological phenomenon. In general, these findings invite the incorporation of assessments of self-identification with groups, to the study of the development of social essentialism.

To conclude, the present studies shed light on a number of aspects regarding social essentialism. First, at a descriptive level, the studies confirm that within the Israeli society, ethnicity stands out as the most inductively powerful human category, above a number of other social categories or personality traits. Second, the studies reveal that ethnicity has a privileged status particularly within a sector of that society, namely, modern religious Zionist Jewish children. This finding underscores the role of cultural input in the emergence of social essentialism. Third, the early emergence of these sectorial differences intimates that cultural input is of a form capable of affecting children's concepts from early on. Fourth, it becomes evident that the distinctive status of ethnicity is only manifest when labels are used, but again this effect is both category and sector specific. Finally, the results implicate the role of self-identification in children's concepts of social categories.

References

Al-Haj, M. (2002). Multiculturalism in deeply divided societies: The Israeli case. *International Journal of Intercultural Relations*, *26*, 169–183.

Allport, G. W. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.

Astuti, R., Solomon, G. E. A., & Carey, S. (2004). Constraints on conceptual development. *Monographs of the Society for Research in Child Development*, *69*(Serial No. 277).

Atran, S. (1990). *Cognitive foundations of natural history*. Cambridge, UK: Cambridge University Press.

Bar-Tal, D. (1996). Development of social categories and stereotypes in early childhood: The case of "the Arab" concept formation, stereotype, and attitudes by Jewish children in Israel. *International Journal of Intercultural Relations*, *20*, 341–370.

Bar-Tal, D., & Teichman, Y. (2005). *Stereotypes and prejudice in conflict: Representation of Arabs in Israeli Jewish society*. Cambridge, UK: Cambridge University Press.

Bigler, R. S., & Liben, L. S. (2007). Developmental intergroup theory: Explaining and reducing children's social stereotyping and prejudice. *Current Directions in Psychological Science*, *16*, 162–166.

Birnbaum, D. (2009). *Children's essentialist beliefs about social categories: The case of Jews and Arabs in Israel*, Unpublished doctoral dissertation, Bar-Ilan University, Ramat-Gan, Israel.

Carey, S. (1995). On the origins of causal understanding. In D. Sperber, D. Premack, & A. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 268–308). Oxford, UK: Clarendon Press.

Central Bureau of Statistics. (2006). *Social Survey—Data 2002–2004*. Jerusalem: The State of Israel Press.

Cimpian, A., Arce, H. C., Markman, E. M., & Dweck, C. S. (2007). Subtle linguistic cues affect children's motivation. *Psychological Science*, *18*, 314–316.

Diesendruck, G., & Haber, L. (2009). God's categories: The effect of religiosity on children's teleological and essentialist beliefs about categories. *Cognition*, *110*, 100–114.

Diesendruck, G., & haLevi, H. (2006). The role of language, appearance, and culture in children's social category based induction. *Child Development*, *77*, 539–553.

Doosje, B., Ellemers, N., & Spears, R. (1995). Perceived intragroup variability as a function of group status and identification. *Journal of Experimental Social Psychology*, *31*, 410–436.

Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong*. New York: Oxford University Press.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. New York: Oxford University Press.

Gelman, S. A., & Heyman, G. D. (1999). Carrot-eaters and creature-believers: The effects of lexicalization on children's inferences about social categories. *Psychological Science*, *10*, 489–493.

Gelman, S. A., Heyman, G. D., & Legare, C. H. (2007). Developmental changes in the coherence of essentialist beliefs about psychological characteristics. *Child Development*, *78*, 757–774.

Gelman, S. A., & Markman, E. M. (1987). Young children's induction from natural kinds: The role of categories and appearances. *Child Development*, *58*, 1532–1541.

Gelman, S. A., Taylor, M. G., & Nguyen, S. (2004). Mother-child conversations about gender: Understanding the acquisition of essentialist beliefs. *Monographs of the Society for Research in Child Development*, *69*(Serial No. 275).

Gil-White, F. J. (2001). Are ethnic groups biological "species" to the human brain? Essentialism in our cognition of some social categories. *Current Anthropology*, *42*, 515–554.

Heyman, G. D., & Diesendruck, G. (2002). The Spanish ser/estar distinction in bilingual children's reasoning about human psychological characteristics. *Developmental Psychology*, *38*, 407–417.

Heyman, G. D., & Gelman, S. A. (2000). Beliefs about the origins of human psychological traits. *Developmental Psychology*, *36*, 663–678.

Hirschfeld, L. A. (1993). Discovering social difference: The role of appearance in the development of racial awareness. *Cognitive Psychology*, *25*, 317–350.

Hirschfeld, L. A. (1996). *Race in the making*. Cambridge, MA: MIT Press.

Kanovsky, M. (2007). Essentialism and folksociology: Ethnicity again. *Journal of Cognition and Culture*, *7*, 241–281.

Levy, S., Levinson, H., & Katz, E. (2002). *Israeli Jews: A portrait.* Jerusalem: Gutman Center Press. (Translated from Hebrew.)

Mahalingam, R. (2003). Essentialism, culture, and power: Representations of social class. *Journal of Social Issues, 59*, 733–749.

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical processing* (pp. 179–195). New York: Cambridge University Press.

Ostrom, T. M., & Sedikides, C. (1992). Out-group homogeneity effects in natural and minimal groups. *Psychological Bulletin, 112*, 536–552.

Patterson, M. M., & Bigler, R. S. (2006). Preschool children's attention to environmental messages about groups: Social categorization and the origins of intergroup bias. *Child Development, 77*, 847–860.

Pitner, R. O., Astor, R. A., Benbenishty, R., Haj-Yahia, M. M., & Zeira, A. (2003). The effects of group stereotypes on adolescents' reasoning about peer retribution. *Child Development, 74*, 413–425.

Sikron, M. (2004). *Demography: Israel's population—Characteristics and trends.* Jerusalem: Carmel Press. (Translated from Hebrew)

Slone, M., Tarrasch, R., & Hallis, D. (2000). Ethnic stereotypic attitudes among Israeli children: Two intervention programs. *Merrill-Palmer Quarterly, 46*, 370–389.

Sloutsky, V. M., Lo, Y., & Fisher, A. V. (2001). How much does a shared name make things similar? Linguistic labels, similarity, and the development of inductive inference. *Child Development, 72*, 1695–1709.

Smooha, S. (1989) *Arabs and Jews in Israel.* Boulder, CO: Westview Press.

Soen, D. (2002). Democracy, the Jewish-Arab cleavage, and tolerance education in Israel. *International Journal of Intercultural Relations, 26*, 215–232.

Sperber, D. (1996). *Explaining culture: A naturalistic approach.* Oxford, UK: Blackwell.

Stendel, O. (1996). *The Arabs in Israel.* Sussex, UK: Sussex Academic Press.

Taylor, M. (1996). The development of children's beliefs about social and biological aspects of gender differences. *Child Development, 67*, 1555–1571.

Teichman, Y. (2001). The development of Israeli children's images of Jews and Arabs and their expression in human figure drawing. *Developmental Psychology, 37*, 749–761.

Waxman, S. R. (1999). The dubbing ceremony revisited: Object naming and categorization in infancy and early childhood. In D. L. Medin & S. Atran (Eds.), *Folkbiology* (pp. 233–284). Cambridge, MA: MIT Press.

Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition, 8*, 223–250.

Child Development, May/June 2010, Volume 81, Number 3, Pages 778–796

Gender, Values, and Occupational Interests Among Children, Adolescents, and Adults

Erica S. Weisgram
University of Wisconsin–Stevens Point

Rebecca S. Bigler
University of Texas at Austin

Lynn S. Liben
The Pennsylvania State University

Work fulfills personal values, perhaps differently for males and females. Explored here was the role values play in shaping occupational interests. Study 1 examined children's, adolescents', and adults' ($N = 313$) occupational values (regarding money, power, family, altruism), occupational interests, and perceptions of values afforded by traditionally masculine and feminine occupations. Results revealed sex differences in occupational values and interests. Furthermore, participants' values predicted their own interests in culturally masculine and feminine occupations. Study 2 used novel jobs and experimentally manipulated prototypical sex of worker and value affordances to disentangle their effects on occupational interests. At all ages, participants' ($N = 240$) occupational interests were affected by the depicted sex of the workers and by the stated value affordances of the jobs.

The distinction between male and female is perhaps the most visible and dramatic subdivision within our species. Within developmental psychology, empirical and theoretical literatures document sex differences in behavioral realms such as play, cognitive skills, peer relations, and educational pathways (Blakemore, Berenbaum, & Liben, 2009; Ruble, Martin, & Berenbaum, 2006). From a lifespan developmental perspective, one of the most striking domains of gender differentiation occurs in the world of work: Substantial sex differences in career interests are evident in preschoolers, elementary-school-age children, adolescents, and adults (Erb, 1983; Herzog, 1982; Schulenberg, Goldstein, & Vondracek, 1991). Although the workforce is, overall, roughly evenly divided by sex, specific jobs remain differentiated. Illustratively, the 2000 U.S. Census (U.S. Department of Commerce, 2005) shows that women hold most jobs in nursing, education, and child care (88%, 81%, and 95%, respectively) whereas men hold most jobs in engineering, computer programming, and physical science (90%, 88%, and 78%, respectively).

The etiology of occupational sex differences has been studied within the fields of psychology, vocational guidance, and counseling psychology for decades, although relatively little work has addressed children's career interests. In a special issue of the *Career Development Quarterly*, McMahon and Watson (2008) called for additional theoretically derived, programmatic research on occupational interests across the lifespan. The work we describe here responds to this call by examining the role of sex-typed occupational values in shaping children's, adolescents', and adults' career

---

This work is based, in part, on the first author's dissertation, which was completed under the direction of the second author and submitted to the University of Texas at Austin. The authors thank the members of the dissertation committee, Jacqueline Woolley, Judith Langlois, William Swann Jr., and Tasha Beretvas, for their insights and suggestions. Many thanks are also extended to Steven Witt, Jamie Garcia, Ellen Hope, Susan Owen, and Amy Weisgram Engstrom for their willingness to invite students to participate in these studies, and to the participants who took part in these studies. We are also appreciative of the undergraduate students in the Gender and Racial Attitudes Lab who assisted in data collection, data entry, and recruitment of subjects, especially Lindsey Fivecoat, Ashley Smith, Masha Sharf, Fran Aucna-Neely, Desiree Sorenson, Marissa Contreras, Lisa Newhouse, John Goosey, and Camille Garvin. Lastly, thanks to Theresa Lesperance for editorial assistance. Portions of this work were presented at the Gender Development Research Conference, April 2006, San Francisco, CA, and at the biennial conference of the Society for Research in Child Development, March 2007, Boston, MA.

Correspondence concerning this article should be addressed to Erica S. Weisgram, University of Wisconsin–Stevens Point, Department of Psychology, 2100 Main Street, Stevens Point, WI 54481. Electronic mail may be sent to eweisgra@uwsp.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8103-0009

interests. Specifically, we tested the hypothesis that some occupational values are endorsed differentially by males and females, and that these differing values contribute to the gender differentiation of occupational interests. In Study 1, we tested the hypothesis that children's, adolescents', and adults' own occupational interests would be related to their career values and of the values afforded by specific occupations. In Study 2, we tested whether children's, adolescents', and adults' occupational interests would be affected by experimental manipulations of jobs' value affordances.

Practical and Theoretical Motivations for Examining the Role of Values in Occupational Goals

The world of work is of practical importance both because it encompasses a large portion of life and because it predicts life quality (e.g., income, mental health, and social relationships; see Crouter, Bumpus, Maguire, & McHale, 1999; Thomas, Benzeval, & Stansfeld, 2007). It is also of theoretical interest for the study of gender differentiation because it involves not only individual development but also societal policies and cultural stereotypes. A recent and highly publicized demonstration of the importance of multiple levels of influence on occupational outcomes comes from an address given by Larry Summers (2005) during his presidency of Harvard University. In his speech, he suggested that the relative dearth of women faculty in engineering and science at prestigious institutions might, in part, be attributed to biological gender differences in cognitive aptitude. This part of his argument became national headline news, contributing to his removal from the Harvard presidency. Although it received far less national attention, Summers also mentioned that another potential explanation for gender differentiation in the work force might lie in sex differences in men's and women's attempts to satisfy life values (e.g., work and family).

Indeed, research has shown that values affect individuals' behaviors and mental processes in various domains, including political beliefs, moral behaviors, and educational attainment (Simpkins, Davis-Kean, & Eccles, 2006). Within the vocational psychology literature, studies have consistently demonstrated that occupational values, typically defined as desires one most wants to be fulfilled within a career (Singer & Stefflre, 1954; Zytowski, 1970), are associated with occupational preferences (Judge & Bretz, 1992; Lindsay & Knox, 1984). Importantly, there is also extensive research demonstrating that some career values are endorsed differentially by males and females (e.g., see Konrad, Ritchie, Lieb, & Corrigall, 2000), although when in the life course such gender differentiation occurs is unknown.

Developmental and Sex Differences in Work Values

Past work has documented that occupational values are under active development during early adolescence. Young adolescents have been found to endorse many values simultaneously (Johnson, 2001; Schulenberg, Vondracek, & Kim, 1993; Weisgram & Bigler, 2006). For example, young adolescents want a job that simultaneously helps people, allows spending time with family, makes a lot of money, and holds influence over others. This set of desires is unrealistic given the values afforded by most occupations. In a longitudinal study, Johnson (2001) found that the proportion of individuals rating multiple values as ''very important'' declined with age; by late adolescence, individuals highly endorsed only a select subset of occupational values. Once adolescents become selective, they generally value money, security, and family time more highly than power and responsibility (Lee, 1984; Thomas, 1986). Little is known, however, about occupational values among children younger than adolescents, despite calls for life-span approaches to understanding career development (Hartung, Porfeli, & Vondracek, 2008).

Although some occupational values are not differentially endorsed by males and females (Rowe & Snizek, 1995), other occupational values are valued more strongly by one gender than the other (Herzog, 1982; Konrad et al., 2000; Schulenberg et al., 1993). Among older adolescents and adults, males typically prefer jobs characterized by high salaries, power or influence over others, opportunities for advancement or achievement, risk taking, challenging tasks, a high level of responsibility, and a high level of prestige (Eccles, 1994; Konrad et al., 2000; Post-Kammer, 1987). Females, in contrast, typically prefer jobs that allow them to work with or help others, develop their knowledge or skills, and spend time with family (Cinamon & Rich, 2002; Eccles, 1994; Konrad et al., 2000; Post-Kammer, 1987). Indeed, Marini, Fan, Finley, and Beutel (1996) reported that gender explained more of the variance in occupational values than race, parental education, maternal employment, community of origin, or religion. The developmental origins of such differences are not, however, well understood.

Effects of Gender-Differentiated Values on Occupational Interests

Based on previous theoretical and empirical work (Eccles, 1994; Eccles et al., 1983; Judge & Bretz, 1992), we hypothesized that sex differences in occupational values may be a driving force in sex differences in occupational interests. Within developmental psychology, Eccles and her colleagues are leading proponents of the notion that values shape individuals' academic and vocational achievement (Eccles, 1987, 1994; Eccles et al., 1983). In their expectancy values model, Eccles and colleagues have conceptualized values in a variety of ways, including the perceived importance or utility of careers, interest or enjoyment of careers, and the cost of choosing one career over another. We focus here on the notion of occupational values as the rewards that are desired from and experienced in work (e.g., money, power; Brown, 2002; Knoop, 1991; Schulenberg et al., 1993), and hypothesize that males and females—at the group level—may seek occupations that fulfill different values. Thus, to the extent that a particular job affords the values prized by one gender more than the other, one would expect to see gender segregation within that particular occupation.

Suggestive evidence for this process has been reported in longitudinal work by Vida and Eccles (2003). They found that females endorse altruistic values to a greater extent than males and that endorsing these values negatively predicts to majoring in math and science at age 20, as well as to having a career in those fields at age 25. Similarly, Weisgram and Bigler (2006) reported that girls who viewed science as affording altruistic values showed greater interest in scientific jobs than girls who did not share this view. Thus, individuals' values and the perceived value affordances of occupations may combine to produce gender differentiation in occupational selections. Furthermore, it seems possible that this effect emerges in childhood (rather than adolescence or adulthood) as a result of children's attention to gender as a social category and their learning of the correlates of maleness and femaleness in the culture (Blakemore et al., 2009; Ruble et al., 2006).

But which jobs afford typically ''masculine'' and typically ''feminine'' values? The answer is unknown. There has been almost no systematic study of the extent to which specific occupations are perceived to afford the satisfaction of various values, or whether children, adolescents, and adults endorse the view that traditionally masculine and feminine occupations—at the group level—afford the fulfillment of different values. One reason for the gap in the literature is the common practice in the career development field of collapsing *across* traditionally masculine and feminine occupations to form career interest subscales (Ji, Lapan, & Tate, 2004; Schulenberg et al., 1991). For example, Holland's (1997) widely used model posits that individuals' occupational interests can be described with reference to six subtypes of occupations (realistic, investigative, artistic, social, enterprising, and conventional). Each subtype includes masculine as well as feminine sex-typed occupation (e.g., ''realistic'' jobs include engineer and dietician). Thus, the question of whether traditionally masculine and feminine jobs are perceived as affording different occupational values is unanswered. It is clear, however, that sex differences characterize adolescents' and adults' interest in many, if not most, of the career subtypes on Holland's and other scales. For example, males score higher on the realistic and investigative Holland subscales, whereas females score higher on the artistic, social, and enterprising subscales (Tracey & Ward, 1998). The issues reviewed above make it apparent that additional research is needed to understand the contributions of both sex typing and values to the development of occupational interests.

A first step in studying the role of values in the gender differentiation of occupational interest is, therefore, documenting children's, adolescents', and adults' views of the values afforded by jobs that are traditionally viewed as masculine versus feminine, and then examining whether individuals' occupational values predict differential levels of interest in these two broad classes of occupations. In addition, research must examine whether the endorsement of occupational values predicts interest in occupations perceived to afford these particular values (e.g., money, power, family, and altruism). Study 1 was designed to fulfill these research goals.

Imagine, however, that—as predicted—masculine versus feminine jobs are found to be perceived by children, adolescents, and adults as affording different values. Why might this be true? It is possible that masculine jobs are characterized by duties and perks that objectively enhance job holders in ways that feminine jobs do not (e.g., the activities engaged in by workers who hold jobs that are culturally viewed as masculine jobs may carry more power). It is also possible, however, that it is not the characteristics of the job itself—but rather the sex of workers who typically hold that job—that shape such perceptions. For example, the mere fact of associating business executives with males may mean that the job of business executive is seen as powerful (because men are seen as powerful). If business executives were instead usually female, the job—even if it had the identical job responsibilities—might then be perceived as being a job that affords family values but *not* be perceived as offering substantial power. Such a process would be consistent with Eagly's (1987) social role theory that posits that gender segregation shapes perceptions of the traits required for occupational roles and, consequently, individuals' occupational interests. Studies that include only measures of participants' interest in existing jobs, although useful for establishing patterns of evidence that are consistent with theoretically derived hypotheses about influence of sex of worker and values on career interests, are unhelpful in disentangling the independent causal effects of these variables. Study 2 was designed to examine this relation by using an experimental methodology to examine the independent roles of (a) perceived value affordances and (b) perceived gender segregation of occupations (i.e., sex of prototypical workers) in occupational interests.

Study 1

In Study 1 we examined children's, adolescents', and adults' self-endorsement of job values (i.e., what values they consider to be important for themselves) and job interests, and whether values and interests differ by participant sex. We focused on four values widely reported in the adult literature to show sex differences: money (sometimes called *salary*, *compensation*, or *economic return*), power (*management* or *authority*), family time, and altruistic values (see Eccles, 1994; Eccles et al., 1983; Post-Kammer, 1987; Weisgram & Bigler, 2006).

Based on past work, we expected females to endorse altruistic and family values more strongly than males and, in contrast, males to endorse money and power values more strongly than females. We expected this pattern would be strongest among adults because children and adolescents were expected to endorse all four values at a generally high rate due to their unrealistic belief that jobs can afford all positive values simultaneously (Johnson, 2001). We also expected to replicate findings from prior studies (Gregg & Dobson, 1980; Liben & Bigler, 2002) that participants would show greater interest in jobs culturally associated with their own sex.

We also examined participants' perceptions of values afforded by a diverse sample of occupations to test our hypothesis that culturally masculine jobs would be rated as higher for fulfilling money and power values than culturally feminine jobs and, conversely, that culturally feminine jobs would be rated higher for fulfilling family and altruistic values than culturally masculine jobs.

Finally, we examined whether individuals' self-endorsements of job values would predict their interests in particular kinds of jobs. To this end, we examined whether individuals' values predicted their relative degree of interest in culturally feminine versus culturally masculine jobs, over and above the variance explained by individuals' biological sex. We expected that participants' personal endorsement of money and power values would predict their interest in masculine occupations, whereas personal endorsement of family and altruistic values would predict their interest in feminine occupations.

Method

Participants

In all, 313 participants were tested. Children (36 boys and 44 girls) ranged from 5 to 10 years ($M = 7.62$, $SD = 1.67$) and included 10 African Americans, 18 Latinos, 50 European Americans, 1 Asian American, and 1 child whose race was unreported. Adolescents (54 boys and 43 girls) ranged from 11 to 17 years ($M = 15.14$, $SD = 1.61$) and included 1 African American, 4 Latinos, 88 European Americans, 2 Asian Americans, and 2 youth whose race was unreported. Children and adolescents were recruited from after-school programs and private schools in the Southwest and Midwest. Adults (73 men and 63 women) ranged from 18 to 23 years ($M = 18.81$, $SD = 1.07$) and included 8 African Americans, 39 Latinos, 61 European Americans, 24 Asian Americans, and 4 individuals whose race was unreported. Adults were recruited from an introductory psychology course at a major university in the Southwest.

Overview of the Procedure

Children were tested individually by the first author or a research assistant, adolescents completed written measures in class, and adults completed measures in small groups. Participants reported their (a) perceptions of the degree to which various occupations afford money, power, family, and altruistic values; (b) occupational values; (c) occupational interests; and (d) demographic characteristics (age, sex, race/ethnicity).

Measures

Perceived value affordances of occupations. To assess beliefs about affordances, children were asked to rate the degree to which each of 40 occupations (drawn from the Occupation, Activity, and Trait scale [OAT]; Liben & Bigler, 2002; described below) "is a job that makes a lot of money/lets people make the rules for others/lets people spend a lot of time with their family/helps people" using a scale ranging from 1 (*not at all*) to 4 (*very much*). Adolescents and adults rated the degree to which each job "provides someone with the opportunity to have a high income/allows someone to be in a position of power/gives someone plenty of time to spend with their family/helps people." Money, power, altruism, and family scores were averaged separately for culturally masculine jobs and for feminine jobs, allowing us to examine whether ratings varied across age groups and allowing us to identify jobs that were viewed as affording money, power, family, or altruism values, and then to examine whether individuals' values predicted to their interest in such occupations. Reliability was acceptable for all subscales within each age group (Cronbach's alphas ranged from .66 to .88, with only two falling below .70). Item analyses revealed similar but not identical factor structures among each age group, with most (but not all) items loading onto the identical factors across age groups.

Endorsement of occupational values. Adolescents' and adults' occupational values were assessed by the 16-item Occupational Values Scale, an instrument developed for young adolescents by Weisgram and Bigler (2006). Children's occupational values were assessed by the Occupational Values Scale for Children, a measure created specifically for this study structured similarly but with developmentally appropriate items (see the Appendix). Participants were asked to indicate how much they would like to have job that incorporates aspects of each of four values (money, power, family, altruism) on a 4-point scale from 1 (*not at all*) to 4 (*very much*).

Cronbach's alphas for the four subscales ranged from .66 to .75 for children, .71 to .91 for adolescents, and .74 to .92 for adults. Item analyses resulted in four factors (money, power, family, and altruism) among adults and adolescents and five factors among children. When a four-factor solution was forced for the children, the same four factors emerged as in the older samples, with items loading identically with an exception of a single item. Responses within each subscale were averaged to create the dependent variables.

Personal occupational interests. Participants were asked to indicate their interest in 40 items drawn from the 80-item long version of the OAT–Personal Measure (Liben & Bigler, 2002). To select the 40 items, we asked an independent sample of adults ($n = 100$; recruited from the undergraduate psychology research pool at a major public university) to rate all 80 items on the four target values and then selected a group of culturally masculine and feminine jobs that would offer as diverse a range of values as possible. For each of the 40 occupations, participants indicated their interest on a scale of 1 (*not at all*) to 4 (*very much*). Cronbach's alphas for the masculine and feminine occupation subscales within each of the age groups ranged between .79 and .94. Item analyses indicated similar but not identical factor structures across age groups. Most (but not all) items loaded on the identical factors across age groups. Summary scores of participants' interest in culturally masculine and feminine occupations were computed by averaging responses to items of each type.

Results

Overview of Analyses

First, we examined participants' perceptions of the values afforded by the target occupations with the goal of determining (a) whether views are shared (or diverge) across age groups, (b) whether culturally masculine and feminine jobs are perceived as affording different values, and (c) whether participants' age and sex were associated with perceptions of the values collectively afforded by culturally masculine versus feminine target occupations. Second, we tested for age and sex differences in participants' personal occupational values and occupational interests. Third, we used hierarchical multiple regression analyses to examine predictors of participants' interest in masculine and feminine occupations. Separate regression analyses were conducted for each sample to examine patterns within each age group.

Perceived Value Affordances of Occupations

To evaluate consistencies across age concerning jobs' value affordances, zero-order correlations were computed for value ratings of each of the 40 occupations, separately for each of the four value categories (money, power, family, and altruism). Agreement was high across age groups, with correlations between children's and adolescents' average value ratings ranging from $r = .79$ to $r = .88$, those between children and adults ranging from $r = .77$ to $r = .90$, and those between adolescents and adults ranging from $r = .95$ to $r = .99$. Given the similar patterns of ratings across age groups, we provide only adults' ratings in Table 1.

To examine possible age or gender differences in perceptions of values afforded by masculine and feminine occupations, each of the four value ratings (i.e., money, power, family, altruism) was analyzed by a separate 2 (participant sex: male, female) × 3 (participant age: children, adolescents, adults) × 2 (sex-type of job: masculine, feminine) mixed-effects-analysis of variance (ANOVA) in which sex-type of job was a within-subjects variable. Means and standard deviations are presented in Table 2. In each case, significant interactions were followed by simple effects analyses (with Bonferonni correction). Results reported as significant here and elsewhere in this article were $p < .05$ or better unless otherwise noted.

Money ratings. The two-way interaction of participant age and sex-type of job was significant, $F(2, 306) = 80.49$. As expected, masculine jobs received higher money ratings than did feminine jobs among all three age groups. The significant interaction stemmed from the fact that the difference between the ratings of masculine and feminine jobs was greater in adults than in either adolescents or children, and greater in adolescents than in children.

Power ratings. The two-way interaction of participant age and sex-type of job was significant, $F(2, 305) = 57.21$. As expected, masculine jobs received higher power ratings than did feminine jobs among all three age groups. Again, the significant interaction stemmed from the fact that the difference between masculine and feminine job ratings was greater in adults than in either adolescents or children, and greater in adolescents than in children.

Family ratings. The two-way interaction of participant sex and sex-type of job was significant, $F(1, 307) = 6.12$. As expected, feminine jobs received higher family values ratings than masculine jobs in both males and females. The significant interaction stemmed from the fact that the difference between ratings of masculine and feminine jobs was greater in females than in males. Also significant was the interaction between participant age and sex-type of job, $F(2, 307) = 26.34$. Feminine jobs were rated as affording family values significantly more than masculine jobs in all three age groups, but the difference was larger in adults than in adolescents or children, and larger in adolescents than in children.

Altruism ratings. The two-way interaction between participant age and sex-type of job was significant, $F(2, 307) = 8.32$. Contrary to expectations, masculine jobs were rated as affording altruistic values significantly more than feminine jobs in all age groups. The significant interaction revealed that the difference between ratings of masculine and feminine jobs was greater in adults than in either adolescents or children, and greater in adolescents than in children.

Occupational Values

Group differences in personal endorsement of occupational values were examined with a 2 (participant sex: male, female) × 3 (participant age: children, adolescents, adults) × 4 (value: money, power, family, altruism) mixed-effects ANOVA with value as a repeated measures factor. Significant interactions were followed by simple effects analyses (with Bonferonni correction). Means and standard deviations are presented in Table 3.

The three-way interaction of participant sex, participant age, and value was significant, $F(6, 918) = 4.12$. Among children, follow-up analyses indicated no significant effects or interactions. Among adolescents, there was a significant Participant Sex × Value interaction, with females endorsing altruistic values significantly more than males. Among adults, the same finding held, and in addition, females endorsed family values significantly more than did males.

Occupational Interests

Group differences in occupational interests were examined by a 2 (participant sex: males, females) × 3 (participant age: children, adolescents, adults) × 2 (sex-type of job: masculine, feminine) mixed-effects ANOVA. Means and standard deviations are presented in Table 3. The three-way interaction of participant sex, participant age, and sex-type of job was significant, $F(2, 307) = 4.13$. Simple effects analyses showed a significant interaction between participant sex and sex-type of job among all three age groups. Among children, girls

Table 1
Adults' Mean Value Ratings (and Standard Deviations) of Occupations Examined in Study 1

Job	Money	Power	Family	Altruism
Masculine jobs				
Airline pilot	3.25 (.69)	3.28 (.72)	1.81 (.72)	3.04 (.56)
Architect	3.49 (.61)	2.96 (.72)	2.52 (.64)	3.01 (.63)
Auto mechanic	2.20 (.62)	2.36 (.77)	2.85 (.54)	3.43 (.57)
Band/orchestra leader	2.29 (.69)	3.24 (.70)	2.83 (.77)	2.52 (.74)
Bus driver	1.51 (.50)	2.59 (.87)	2.93 (.75)	3.26 (.59)
Chemist	3.20 (.62)	2.68 (.73)	2.20 (.74)	3.14 (.72)
Computer builder	3.17 (.68)	2.49 (.74)	2.67 (.73)	3.04 (.77)
Dentist	3.80 (.42)	3.28 (.67)	2.48 (.78)	3.75 (.44)
Doctor	3.93 (.26)	3.59 (.59)	1.89 (.78)	3.93 (.25)
Farmer	1.97 (.76)	2.16 (.82)	2.91 (.84)	3.25 (.68)
Firefighter	2.19 (.64)	2.94 (.82)	2.23 (.73)	3.91 (.29)
Football broadcaster	2.99 (.77)	2.15 (.78)	3.06 (.81)	2.11 (.75)
Janitor	1.35 (.48)	1.60 (.58)	2.81 (.72)	3.07 (.74)
Lawyer	3.88 (.37)	3.57 (.57)	1.96 (.74)	3.51 (.66)
Parking lot attendant	1.39 (.50)	1.88 (.76)	3.29 (.68)	2.46 (.75)
Police officer	2.45 (.65)	3.70 (.49)	2.18 (.70)	3.77 (.49)
President of the United States	3.72 (.50)	3.96 (.21)	1.78 (.81)	3.49 (.65)
Professional athlete	3.93 (.31)	3.06 (.88)	1.92 (.79)	1.88 (.68)
Refrigerator salesperson	1.79 (.59)	1.76 (.56)	3.15 (.65)	2.24 (.76)
Scientist	3.10 (.66)	2.75 (.73)	2.16 (.74)	3.26 (.73)
Shoe repairer	1.63 (.52)	1.79 (.62)	3.12 (.60)	3.00 (.62)
Supreme Court judge	3.67 (.49)	3.95 (.22)	2.06 (.90)	3.21 (.81)
Telephone installer	1.82 (.58)	1.87 (.57)	3.04 (.58)	2.88 (.71)
Feminine occupations				
Babysitter/nanny	1.70 (.61)	2.69 (.99)	3.01 (.77)	3.56 (.54)
Cheerleader	1.65 (.64)	1.90 (.79)	3.20 (.79)	1.89 (.76)
Elementary school teacher	1.90 (.59)	3.20 (.80)	3.14 (.68)	3.85 (.42)
Florist	1.73 (.55)	1.71 (.60)	3.31 (.65)	2.24 (.75)
Hair stylist	2.10 (.72)	2.09 (.85)	3.22 (.68)	2.68 (.74)
Jewelry maker	2.57 (.83)	1.91 (.70)	3.04 (.70)	2.11 (.67)
Librarian	1.86 (.53)	2.18 (.69)	3.23 (.54)	3.13 (.71)
Nurse	2.82 (.57)	2.73 (.72)	2.35 (.70)	3.85 (.36)
Perfume salesperson	1.66 (.61)	1.66 (.59)	3.33 (.69)	2.02 (.80)
Secretary	2.07 (.51)	2.05 (.59)	3.04 (.54)	3.11 (.65)
Social worker	2.08 (.61)	2.47 (.85)	2.76 (.61)	3.66 (.62)
Supermarket clerk	1.51 (.52)	1.75 (.64)	3.27 (.58)	2.75 (.73)
Telephone operator	1.68 (.52)	1.94 (.64)	2.99 (.63)	2.93 (.71)
Neutral occupations				
Artist	2.47 (.91)	2.32 (.93)	3.10 (.74)	2.37 (.75)
Baker	1.97 (.55)	1.96 (.69)	3.00 (.63)	2.69 (.64)
Elevator operator	1.40 (.49)	2.04 (.83)	3.19 (.69)	2.51 (.83)
Writer	2.74 (.83)	2.51 (.83)	3.11 (.78)	2.44 (.75)

Note. Value ratings range from 1 (*low*) to 4 (*high*).

were significantly more interested in culturally feminine occupations than were boys; among adolescents, males were significantly more interested in culturally masculine occupations than were females; and among adults, males were significantly more interested in culturally masculine occupations—and less interested in culturally feminine occupations—than were females.

Predictors of Interest in Culturally Masculine and Feminine Occupations

To examine the predictors of individuals' interests in culturally masculine and feminine occupations, hierarchical multiple regression analyses were conducted in which participant age, participant sex, endorsement of each of the four

Table 2
Value Ratings of Masculine and Feminine Occupations

	Money		Power		Family		Altruism	
	Masculine	Feminine	Masculine	Feminine	Masculine	Feminine	Masculine	Feminine
Children								
Males	3.15 (.42)	2.88 (.52)	2.81 (.58)	2.64 (.65)	2.68 (.60)	2.85 (.68)	3.26 (.48)	3.00 (.66)
Females	3.14 (.42)	2.93 (.56)	2.81 (.50)	2.71 (.60)	2.59 (.42)	2.86 (.43)	3.21 (.41)	2.97 (.52)
Combined	3.15 (.42)	2.91 (.54)	2.81 (.54)	2.68 (.62)	2.63 (.51)	2.86 (.56)	3.23 (.44)	2.98 (.59)
Adolescents								
Males	2.91 (.29)	2.26 (.50)	2.83 (.33)	2.30 (.50)	2.57 (.32)	2.92 (.38)	3.01 (.33)	2.68 (.43)
Females	2.84 (.26)	2.19 (.35)	2.90 (.37)	2.32 (.57)	2.46 (.32)	2.94 (.39)	3.07 (.39)	2.71 (.46)
Combined	2.87 (.27)	2.22 (.42)	2.87 (.35)	2.31 (.54)	2.51 (.32)	2.93 (.39)	3.04 (.36)	2.69 (.45)
Adults								
Males	2.71 (.21)	1.93 (.28)	2.70 (.32)	2.08 (.41)	2.51 (.30)	3.04 (.37)	3.03 (.30)	2.82 (.39)
Females	2.74 (.24)	1.97 (.28)	2.84 (.32)	2.28 (.45)	2.52 (.33)	3.09 (.44)	3.17 (.38)	3.01 (.39)
Combined	2.72 (.23)	1.95 (.28)	2.76 (.33)	2.17 (.44)	2.51 (.32)	3.07 (.40)	3.09 (.35)	2.90 (.39)

Note. Value ratings range from 1 (*low*) to 4 (*high*).

Table 3
Value Endorsement and Interest in Familiar Occupations

	Occupational values				Interest	
	Money	Power	Family	Altruism	Masculine	Feminine
Children						
Males	3.79 (.50)	3.17 (.90)	3.49 (.60)	3.58 (.52)	2.38 (.94)	2.06 (1.01)
Females	3.56 (.54)	3.06 (.67)	3.57 (.54)	3.61 (.48)	2.14 (.62)	2.52 (.63)
Adolescents						
Males	3.63 (.57)	3.37 (.60)	3.63 (.46)	3.26 (.60)	2.11 (.46)	1.44 (.47)
Females	3.58 (.55)	3.21 (.67)	3.67 (.48)	3.62 (.46)	1.71 (.32)	1.61 (.41)
Adults						
Males	3.47 (.66)	3.25 (.58)	3.42 (.55)	3.46 (.54)	2.06 (.39)	1.36 (.30)
Females	3.38 (.62)	3.17 (.58)	3.73 (.39)	3.76 (.34)	1.59 (.34)	1.78 (.41)

Note. Value scales for all samples range from 1 to 4, with higher scores indicating stronger endorsement of the value. Interest scales range from 1 to 4, with higher scores indicating greater personal interest in the occupation.

occupational values, and the various interactions between age, sex, and the endorsement of occupational values served as predictors of participants' interest in masculine and feminine occupations separately. For interest in masculine occupations, this model was found to be significant, $F(28, 280) = 5.18$, $R^2 = .34$. Importantly, age produced interaction terms with each of the values. For interest in feminine occupations, this model was also significant, $F(28, 280) = 8.53$, $R^2 = .46$. Again, age was involved in significant interactions with each of the values. To clarify these age interactions, hierarchical multiple regression analyses were conducted within each of the age groups (i.e., children, adolescents, and adults). In each model, predictor variables included participant sex, participant age in years, and participants' endorsement of each of the four occupational values. Results of these regressions are summarized below and detailed in Table 4.

For each regression, participant sex and age were entered at the first step of the model for all three age groups. At Step 2, endorsement of each occupational value was entered. Tests of multicollinearity (i.e., tests of minimum tolerance) among constructs indicated that collinearity was at an acceptable level for each analysis. Table 4 presents beta values, standard errors, and standardized betas for regressions predicting interest in masculine and feminine occupations.

Children. For masculine occupations, the overall model was significant in predicting children's

Table 4
Predictors of Interest in Masculine and Feminine Occupations

	Masculine				Feminine			
	B	*SE*(*B*)	β	R^2	*B*	*SE*(*B*)	β	R^2
Children (*N* = 80)								
Step 1								
Sex	−.173	.160	−.111	.22*	.538*	.146	.313	.46
Age	−.218*	.049	−.463		−.317*	.043	−.623	
Step 2								
Sex	−.147	.156	−.094	.34*	.556*	.140	.323	.55*
Age	−.194*	.049	−.403		−.293*	.042	−.575	
Money	.103	.163	.070		.071	.148	.043	
Power	.219*	.109	.220		−.241*	.100	−.216	
Family	.144	.148	.104		.094	.132	.061	
Altruism	.240	.160	.153		.280*	.145	.160	
Adolescents (*N* = 97)								
Step 1								
Sex	−.404*	.079	−.464	.22*	.160	.088	.181	.08
Age	−.022	.023	−.102		−.054	.027	−.198	
Step 2								
Sex	−.426*	.084	−.489	.27	.166	.094	.131	.14
Age	−.022	.023	−.085		−.055*	.028	−.199	
Money	.095	.086	.122		.155	.098	.197	
Power	.026	.074	.038		−.164*	.082	−.241	
Family	.075	.090	.081		.064	.101	.068	
Altruism	.079	.078	.101		.070	.087	.089	
Adults (*N* = 136)								
Step 1								
Sex	−.454*	.065	−.524	.30*	.417*	.061	.506	.25*
Age	.033	.030	.082		.032	.029	.083	
Step 2								
Sex	−.502*	.069	−.579	.37*	.325*	.066	.395	.35*
Age	.019	.029	.048		.025	.028	.064	
Money	.095	.058	.140		−.006	.056	−.009	
Power	.142*	.067	.180		−.198*	.065	−.279	
Family	.001	.067	.001		.127†	.065	.155	
Altruism	.167*	.075	.183		.156*	.073	.181	

†$p < .10$. *$p < .05$.

interest, $F(6, 72) = 6.65$, $R^2 = .34$, adjusted $R^2 = .28$. Age was a negative predictor of interest in masculine occupations. In addition, the endorsement of power values positively predicted interest in masculine occupations, after accounting for participant sex and age. For feminine occupations, the overall model was significant in predicting children's interest, $F(6, 72) = 15.27$, $R^2 = .56$, adjusted $R^2 = .52$. Sex of participant was a significant predictor of interest in feminine occupations, with girls showing greater interest than boys. Age was a negative predictor of interest in feminine occupations. After accounting for the effects of participant sex and age, altruistic values positively predicted interest in feminine occupations.

Adolescents. For masculine occupations, the overall model was significant in predicting adolescents' interest, $F(6, 90) = 5.53$, $R^2 = .27$, adjusted $R^2 = .22$. Sex of participant was the only significant predictor of interest in masculine occupations, with males showing greater interest than females. For feminine occupations, the overall model was significant in predicting adolescents' interest, $F(6, 90) = 2.20$, $R^2 = .13$, adjusted $R^2 = .08$. Age was a negative predictor of interest in feminine occupations. After accounting for participant sex and age, the endorsement of power values negatively predicted interest in feminine occupations.

Adults. For masculine occupations, the overall model was significant in predicting adults' interest,

$F(6, 130) = 12.88$, $R^2 = .37$, adjusted $R^2 = .34$. Sex of participant was a significant predictor of interest in masculine occupations, with men showing greater interest than women. In addition, the endorsement of altruistic and power values positively predicted interest in masculine occupations. For feminine occupations, the overall model was significant in predicting adults' interest, $F(6, 130) = 11.56$, $R^2 = .35$, adjusted $R^2 = .32$. Participant sex predicted interest in feminine occupations, with women showing greater interest than men. In addition, the endorsement of altruistic and family values positively predicted interest. The endorsement of power values negatively predicted interest in feminine occupations.

Discussion

The primary purpose of Study 1 was to examine individuals' occupational interests in relation to sex, age, and own occupational values. Participants rated first, how much they believed various occupations satisfy money, power, family, and altruistic values; and second, their own occupational values and occupational interests.

Interestingly, the data indicate a very high degree of agreement among the three age samples concerning values afforded by specific jobs. Individuals' perceptions of the values collectively afforded by culturally masculine and feminine occupations were also informative. As hypothesized, masculine jobs were perceived as affording money and power values more than feminine jobs. Feminine jobs, in contrast, were perceived as affording family values to a greater extent than masculine jobs. Interestingly, and contrary to expectations, masculine occupations were perceived by all groups as affording altruistic values to a greater extent than feminine occupations. Although children perceived masculine occupations as affording altruistic values more than feminine occupations, the altruism ratings of feminine occupations were elevated relative to the money, power, and family ratings of these same occupations. In other words, traditionally feminine occupations were perceived as less useful than traditionally masculine jobs for fulfilling most values and were best at fulfilling altruistic values (albeit still less useful for this purpose than masculine jobs).

Consistent with prior research, participants' own occupational values showed significant sex differences. As predicted, in both adolescent and adult groups, altruistic and family values were rated more highly by females than males. In these age groups, ratings for power and money were also in the predicted direction (i.e., males showing greater endorsement than females for money and power values), but these differences were not significant. This pattern—females endorsing traditionally masculine as well as feminine traits and interests, and men avoiding traditionally feminine traits and interests—characterizes much of the literature on sex differences (Konrad et al., 2000; Ryckman & Houston, 2003). Among children, however, there were no sex differences in occupational values. Instead, children strongly endorsed all four values (money, power, family, and altruism). Perhaps it is during middle childhood that children become aware that a single occupation cannot realistically afford all positive values simultaneously, and this realization, in turn, prompts the development of a more strongly prioritized (and sex-differentiated) set of occupational values.

With respect to occupational interests, predicted sex differences in masculine and feminine occupations occurred. Specifically, among children and adults (but not adolescents), females were significantly more interested in feminine occupations than were males. Among adolescents and adults (but not children), males were more interested in masculine occupations than were females. These data are thus highly consistent with past work on occupational interests in indicating that, with the exception of a brief period in late childhood and early adolescence, females are more interested in feminine occupations than males despite the fact that females themselves rate feminine jobs as lower in salary, power, and altruism (Liben, Bigler, & Krogh, 2001). This finding is intriguing and unexplained in the literature. Low self-esteem, academic competence, and self-efficacy have been reported among girls, especially within the domain of mathematics, and could potentially explain their low aspirations (Watt & Eccles, 2008). The data are, however, not entirely consistent with the idea that a lack of confidence within traditionally masculine domains lowers girls' occupational aspirations. Girls receive better grades in all academic subjects than boys and sometimes report high levels of academic competence, self-efficacy, and aspirations within masculine domains such as science (Blakemore et al., 2009; Schulenberg et al., 1991). Furthermore, it is difficult, if not impossible, on the basis of existing data to establish the direction of causal relations (if present) among interest, self-efficacy, and expectations. Such findings leave open the possibility that girls prefer feminine occupations—perhaps because such jobs afford the fulfillment of their values—rather than expect to be unqualified for masculine occupations.

Our primary interest—the role of occupational values in shaping individuals' occupational interest—was examined using regression models. Individuals' endorsement of values was a significant predictor of their occupational interests in several cases. The endorsement of power values positively predicted interest in masculine occupations among children and adults, and negatively predicted interest in feminine occupations among all three samples (children, adolescents, and adults). Thus, the desire for a job that affords power was a clear factor in differentiating interest in masculine versus feminine occupations. In addition, the endorsement of altruistic values positively predicted interest in masculine occupations among adults. Given that masculine occupations included in this research were rated as high on the altruism scale, this finding is not entirely surprising. This same finding did not hold true, however, within our younger samples and, indeed, the reverse held among children. That is, children who endorsed altruistic values were more (rather than less) interested in feminine occupations than children who did not endorse altruistic values. These data are interesting in that they suggest that the sex typing of occupations, gender, and altruistic values may be conceived of, and related to each other, in ways that differ over the life course.

Overall, results show sex differences in occupational values in adults and adolescents, and sex differences in occupational interests at all three ages. Data also suggest that values (especially power and altruism) are connected to occupational interests at all ages.

Although the findings discussed above are generally consistent with the hypotheses we derived from earlier literature, the relevant data are limited in a number of important ways. One important limitation—which the data from Study 1 themselves confirm—is that sex typing of occupations is, overall, confounded with the values associated with occupations. Thus, the data in Study 1 do not allow us to discern, for example, whether males are more interested than females in masculine occupations because these "masculine" occupations afford male-typed values (e.g., money, power) or instead because these occupations are dominated by male workers. Indeed, it is impossible to disentangle the effects of these two factors (sex typing and values) by collecting data concerning actual occupations because these two qualities are confounded in the job structure of our culture. A second limitation is that the correlational nature of Study 1 does not allow us to disentangle the effects of the values afforded by an occupation from other qualities of the job (e.g., sex of workers, specific duties). Thus, experimental research is necessary to determine whether occupational values can have a casual effect on individuals' occupational interests. Because we deemed it unethical to experimentally manipulate individuals' values, we sought to examine the effects of manipulating the value affordances of particular jobs. To do so, we introduced participants to novel occupations whose characteristics would therefore be unknown and whose critical qualities (i.e., the values the jobs afford, the prototypical sex of workers in those jobs) could be experimentally manipulated.

Study 2

In Study 2 we conducted an experimental test of individuals' occupational interests as a function of the separate contributions of value affordances (i.e., the values afforded by a given job) and sex of worker (i.e., the percentage of males and females workers in a given job). Previous research has documented that individuals show higher levels of interest in occupations that are performed by same-sex than other-sex individuals (Heilman, 1979; Touhey, 1974). Thus, it was important to manipulate the role of values independently of sex of worker. To do so, we used an adaptation and extension of Liben et al.'s (2001) novel job paradigm. Participants were introduced to eight novel jobs whose value affordances were manipulated between subjects. That is, each job was described as satisfying one of four values: money, power, family, and altruism. We predicted that males would be more interested in jobs that were depicted as affording money and power than would females, whereas females would be more interested in jobs that were depicted as affording family and altruistic values than would males. We expected such effects to emerge by adolescence, and possibly by late childhood.

To disentangle the effect of value affordances from the effect of sex typing of occupations, we also manipulated the sex of workers depicted in each novel occupation: Half the novel jobs were depicted with female workers and half were depicted with male workers. Based on previous research (Plost & Rosen, 1974), we expected males to be significantly more interested in jobs depicted with male than female workers and, in contrast, females to be more interested in jobs depicted with female than male workers. Each of the novel jobs appeared in each combination of depicted sex of worker and value (i.e., stimuli were completely counterbalanced). Thus, the effect of sex of worker, value affordance, and their interaction could be determined independent of the specific job portrayed.

Method

Participants

Participants included 240 individuals. Children (40 boys and 40 girls; all European Americans) between 5 and 10 years ($M = 7.53$, $SD = 1.53$) were recruited from after school programs in a mid-sized city in the Southwest. Adolescents (40 boys and 40 girls: 1 African American, 1 Hispanic American, 1 Asian American, 77 European Americans) between 11 and 17 years ($M = 14.06$, $SD = 1.13$) were recruited from summer programs and a private school in the Midwest. Adults (40 men and 40 women: 2 African Americans, 6 Hispanic Americans, 4 Asian Americans, 68 European Americans) were undergraduates attending a large university in the Southwest (ages 18–25; $M = 18.81$, $SD = 1.28$) who participated in exchange for course credit.

Overview of Procedure

Participants received a brief introduction to eight novel occupations. They then heard a detailed description of each job and completed measures of their perception of values represented by, perception of sex of workers in, their own interest in, and stereotyping of the job. Participants had visual access to depictions of the novel jobs while completing these measures.

Experimental Manipulation

As described above, Study 2 used 8 of the 12 novel occupations developed by Liben et al. (2001), selected on the basis of pilot data showing that these job descriptions could be modified to convey different value affordances. For example, the job of benster was first described as ''a person who studies deer. They collect information about where deer live, how much food they need, how to keep them off the highways, and how pollution might be hurting them.'' The four value affordances were presented by continuing the job description in one of the following four ways: ''Bensters make a lot of money studying deer. Because they get special training to become experts on deer, they then get paid a lot of money to do their job'' (money); ''Bensters have a lot of power because they train other deer specialists and tell them what to do. Bensters also make very important decisions and have a lot of responsibility'' (power); ''Bensters only work when deer are out so they have lots of time to spend with their children'' (family); or ''Bensters help deer by keeping them healthy and keeping them safe. Bensters also help deer by feeding them and taking care of them'' (altruism).

During training, participants were told they would learn about eight new occupations. For each occupation, a job description was provided, accompanied by a poster showing either six men or six women performing the occupation. Specific jobs were counterbalanced across gender and value condition to ensure that posttest responses reflected only the experimental manipulations and not effects of specific jobs. Posters remained in view while the four measures were given.

Measures

First, to examine whether participants had attended to the sex of worker manipulation, for each job, participants were asked: ''Who is usually a _____ [e.g., benster]?'' Response options included: ''men,'' ''women,'' or ''both men and women.'' Second, to test whether participants had attended to the values manipulation, participants were asked to how much each occupation afforded money, power, family, and altruistic values. For example, participants were asked: ''How much money do bensters make?'' Response options included: ''none,'' ''a little,'' ''some,'' and ''a lot.'' Third, to assess stereotyping of novel occupations, participants were asked: ''Who should be a(n) __________?'' Following Liben and Bigler (2002), response options included ''only women,'' ''only men,'' or ''both men and women.'' Finally, to assess interest in the novel occupations, participants were asked: ''How much would you like to be a(n) ________?'' Participants responded using a 4-point scale ranging from *not at all* to *very much*.

Results

Knowledge of Sex of Worker

To evaluate participants' awareness of the depicted sex of worker, we tallied the number of correct responses to the question: ''Who usually is a

_____?'' Participants correctly identified the workers' gender at a high rate (98%, 86%, and 91% for children, adolescents, and adults, respectively). Participants who answered incorrectly never reversed the depicted sex of worker but instead responded that ''both men and women'' usually perform the occupation.

Perceptions of Values Afforded by Occupations

To ascertain if we successfully manipulated participants' perceptions of the values afforded by the novel occupations, a 3 (age of participant: child, adolescent, adult) × 2 (sex of worker: male, female) × 4 (manipulated value: money, power, family, altruism) ANOVA was conducted using each of the four value ratings of each job (i.e., money, power, family, altruism) as a dependent variable. Significant main effects were followed by simple contrasts among means; means and standard deviations are presented in Table 5.

Money. For money ratings, a significant interaction between age group and manipulated value was found, $F(6, 708) = 11.69$. Within each age a significant effect of manipulated value was evident: children, $F(3, 237) = 64.64$; adolescents, $F(3, 237) = 237.47$; and adults, $F(3, 237) = 368.54$. Within age, follow-up tests revealed participants' money ratings of jobs depicted as high in salary were—as expected—significantly greater than money ratings of jobs depicted with other values (i.e., power, family, altruism). Other patterns differed by age: In adolescents and adults (but not children) jobs depicted as high in power were rated as significantly higher in affording money values than were jobs depicted as high in family or altruism.

Power. For power ratings, a significant interaction between age group and manipulated value was found, $F(6, 711) = 8.31$. Within age groups, a significant effect of manipulated value was found: children, $F(3, 237) = 64.21$; adolescents, $F(3, 237) = 155.21$; and adults, $F(3, 237) = 275.43$. Follow-up tests revealed that—as expected—participants' ratings of power were higher for jobs depicted as high in power than for jobs depicting other values (i.e., money, family, altruism). Other contrasts, however, differed by age: Among adolescents and adults (but not children), jobs depicted as high in money were rated as significantly higher in affording power values than were jobs depicted as high in family or altruism.

Family. For family ratings, a significant interaction between age group and manipulated value was found, $F(6, 708) = 4.13$. Within age, a significant effect of manipulated value was evident: children, $F(3, 237) = 117.85$; adolescents, $F(3, 237) = 307.75$; and adults, $F(3, 237) = 334.10$. Within age, follow-up tests showed that family ratings of jobs depicted as allowing time with family were —as expected—significantly greater than family ratings of jobs depicted with other values. Other

Table 5
Mean Value Ratings (and Standard Deviations) by Manipulated Value and Sex of Worker in Study 2

	Manipulated value				Sex of worker	
	Money	Power	Family	Altruism	Males	Females
Children's value ratings						
Money	3.98 (.12)	3.13 (.78)	3.01 (.76)	3.06 (.80)	3.29 (.54)	3.29 (.54)
Power	2.99 (.88)	3.98 (.14)	2.82 (.89)	2.91 (.99)	3.17 (.62)	3.18 (.64)
Family	2.63 (.78)	2.59 (.78)	3.93 (.20)	2.63 (.81)	2.98 (.54)	2.91 (.54)
Altruism	3.36 (.67)	3.25 (.88)	3.26 (.78)	3.91 (.23)	3.44 (.60)	3.45 (.49)
Adolescents' value ratings						
Money	3.93 (.33)	2.86 (.53)	2.46 (.60)	2.53 (.46)	2.96 (.35)	2.93 (.44)
Power	2.47 (.71)	3.81 (.41)	2.17 (.65)	2.27 (.75)	2.67 (.46)	2.68 (.50)
Family	2.34 (.56)	2.29 (.50)	3.93 (.25)	2.66 (.50)	2.77 (.39)	2.84 (.33)
Altruism	2.64 (.84)	2.76 (.81)	2.64 (.71)	3.67 (.52)	2.94 (.52)	2.92 (.58)
Adults' value ratings						
Money	3.97 (.15)	2.74 (.44)	2.35 (.45)	2.36 (.48)	2.88 (.30)	2.83 (.31)
Power	2.28 (.64)	3.86 (.44)	1.99 (.58)	2.03 (.60)	2.56 (.45)	2.55 (.45)
Family	2.39 (.60)	2.46 (.45)	3.98 (.16)	2.70 (.48)	2.82 (.34)	2.95 (.37)
Altruism	2.33 (.57)	2.36 (.67)	2.43 (.63)	3.73 (.48)	2.73 (.38)	2.69 (.47)

Note. Value ratings range from 1 (*low*) to 4 (*high*).

patterns differed by age: For adolescents and adults (but not children) jobs depicted as high in altruism were rated as significantly higher in affording family values than jobs depicted as high in money or power.

A significant interaction between age group and sex of worker was also found, $F(2, 236) = 7.97$. Within age, a significant effect of sex of worker was present in adolescents $F(1, 79) = 5.01$, and adults, $F(1, 79) = 16.94$, but not children, $F(1, 79) = 3.05$, $p = .09$. Among adolescents and adults (but not children), jobs depicted with female workers were perceived as affording family values significantly more than jobs depicted with male workers.

Altruism. For altruism ratings, a significant interaction of age group and manipulated value was found, $F(6, 708) = 9.36$. Within age, a significant effect of manipulated value was found: children, $F(3, 237) = 27.08$; adolescents, $F(3, 237) = 49.77$; adults, $F(3, 237) = 115.62$. Within age, altruism ratings of jobs depicted as helping others were—as expected—significantly greater than altruism ratings of jobs depicting other values.

Egalitarian Attitudes Toward Novel Occupations

A mixed-effects ANOVA was performed on participants' egalitarian attitudes toward novel occupations. Between-subject variables were sex of participant (male, female) and age group (children, adolescents, adults). Within-subject variables were sex of depicted workers (male, female) and manipulated value (money, power, family, altruism). Dependent measures were proportions of egalitarian responses (i.e., "both men and women" should do the job).

Results indicated a significant three-way interaction of age group, sex of depicted worker, and manipulated value, $F(6, 699) = 2.40$. Simple effects analyses indicated a marginally significant interaction between sex of depicted worker and manipulated value for children, $F(3, 237) = 2.61$, $p = .06$, but not for adolescents or adults. Follow-up tests indicated that children held more egalitarian attitudes toward jobs depicted as high in salary when male workers were depicted ($M = .59$, $SD = .50$) than when female workers were depicted ($M = .38$, $SD = .49$). This trend was not found in jobs depicted as high in power, family, or altruism. Among adolescents and adults, however, a significant main effect of sex of depicted worker was present: for adolescents, $F(1, 78) = 11.20$, and for adults, $F(1, 79) = 12.70$. Across all manipulated values, adolescents and adults held more egalitarian attitudes toward jobs depicted with male workers ($M = .92$, $SD = .18$ and $M = .96$, $SD = .15$, respectively) than with female workers ($M = .84$, $SD = .25$ and $M = .88$, $SD = .24$, respectively).

Interest in Novel Occupations

A mixed-effects ANOVA was performed with between-subjects variables of sex of participant (male, female) and age (children, adolescents, adults) and within-subjects variables of sex of depicted workers (males, females) and manipulated value (money, power, family, altruism). Following interactions, simple effects analyses were performed to explore group differences on the dependent variable. Means and standard deviations are presented in Table 6.

Results indicated a significant three-way interaction of age group, sex of participant, and sex of workers, $F(2, 233) = 9.92$. To explore this interaction, separate mixed-effects ANOVAs were performed for each age group with sex of participant (male, female) as the between-subjects variable and sex of depicted workers (males, females) as the within-subjects variable. Within each age group, a significant interaction between sex of depicted worker and sex of participant was found: children, $F(1, 78) = 57.04$; adolescents, $F(1, 78) = 16.34$; and adults, $F(1, 78) = 13.43$. Among children and adolescents, participants showed greater interest in jobs depicted with members of their own sex. Among adults, jobs depicted with male workers were significantly more attractive to men than women; jobs depicted with female workers were not differentially attractive in relation to participant sex.

A significant interaction between sex of participant and manipulated value was also found, $F(3, 699) = 2.90$. Follow-up analyses indicated that males, across all ages, were significantly more interested in jobs when they were depicted as high in money than when the identical jobs were depicted as high in power, family, or altruism. Like males, females were significantly more interested in jobs depicted as high in money than in the identical jobs depicted as high in power and family; unlike males, however, their interest did not differ between jobs depicted as high in money and those depicted as high in altruism. In addition, males (but not females) were significantly more interested in jobs that were depicted as high in power and family than in identical jobs that were depicted as high in altruism.

Table 6
Means (and Standard Deviations) for Interest in Novel Occupations Manipulated Value and Sex of Worker in Study 2

	Manipulated value				Sex of worker	
	Money	Power	Family	Altruism	Males	Females
Children						
Boys	2.29 (.87)	2.08 (.92)	2.24 (.97)	2.01 (.86)	2.51 (.78)	1.79 (.73)
Girls	2.41 (.98)	2.13 (.77)	2.24 (.84)	2.34 (.75)	2.06 (.63)	2.50 (.79)
Combined	2.35 (.92)	2.10 (.84)	2.24 (.90)	2.18 (.82)	2.28 (.74)	2.15 (.84)
Adolescents						
Boys	1.90 (.66)	1.76 (.58)	1.61 (.58)	1.61 (.58)	1.90 (.52)	1.54 (.46)
Girls	1.96 (.88)	1.80 (.81)	1.71 (.75)	1.90 (.82)	1.78 (.66)	1.91 (.71)
Combined	1.93 (.77)	1.78 (.70)	1.66 (.67)	1.76 (.72)	1.84 (.59)	1.73 (.62)
Adults						
Men	2.21 (.75)	1.78 (.81)	1.81 (.81)	1.59 (.66)	2.04 (.63)	1.66 (.67)
Women	1.93 (.66)	1.70 (.77)	1.69 (.72)	1.71 (.60)	1.74 (.53)	1.78 (.57)
Combined	2.10 (.71)	1.74 (.79)	1.75 (.77)	1.65 (.63)	1.89 (.60)	1.72 (.62)

Note. Scores range from 1 to 4, with higher scores indicting greater personal interest in performing the occupation.

Discussion

The primary purpose of Study 2 was to determine whether children's, adolescents', and adults' interests in occupations are affected, independently, by (a) the values afforded by occupations and (b) the sex of workers employed within occupations, apart from whatever links exist between these two factors. To accomplish this goal, participants were presented with novel jobs that varied with respect to the depiction of the sex of workers and the description of the values that are afforded by the occupations. Dependent measures included participants' ratings of the values afforded by the jobs, the degree to which participants held egalitarian attitudes about the jobs, and the extent to which participants were personally interested in the jobs.

As a first step, we conducted a manipulation check to make sure that participants had acquired intended knowledge of the workers' gender and of the jobs' value affordances. With respect to gender, participants were almost at ceiling in correctly reporting of the sex of workers that had been depicted in the manipulation. With respect to value affordances, data indicated that the experimental manipulation functioned as intended. That is, participants rated each occupation as affording the value that had been depicted in the manipulation more strongly than any of the other three values.

Unexpectedly, the data for the manipulation check revealed interesting developmental changes in the interrelations among different job values. Among adolescents and adults, there were a number of links among the value ratings of occupations. Specifically, among these older participants, ratings of the degree to which jobs afford money were affected by the manipulation of other values: Their money ratings were significantly higher when told that a job was high in power than when told that the identical job was high in family and altruism. Similarly, these older participants' ratings of the degree to which jobs afford power were affected by the manipulation of other values; their power ratings were significantly higher when told that a job was high in money than when told that the identical job was high in family and altruism. Finally, older participants' ratings of the degree to which jobs afford family values were affected by other values; their family ratings were significantly higher when told that a job was high in altruism than when told that the identical job was high in money or power. Collectively, these findings indicate that adolescents and adults (but not children) have developed an occupational schema that includes links between money and power and between altruism and family values. These data may reflect the association between these value pairs in jobs in actual occupations, and these might be mediated through associations with masculine and feminine jobs. The emergence of such links would be interesting to study in its own right.

As discussed in presenting the rationale for Study 2, one possible basis for interpreting jobs as affording different job values is the link between a job and the sex of workers typically employed in the role, or the sex-type of the occupation. Supporting the hypothesized influence of worker sex on judgments of value affordance was the finding that, among adolescents and adults (but not children), sex of worker affected the perceived value affordances of jobs. Specifically, jobs depicted with female workers were perceived as affording more time with one's family than did the identical jobs depicted with male workers.

Theorists have argued that sex of worker is an important variable because a preponderance of one sex within fields leads to the stereotyping of occupations as appropriate for only one gender (Bussey & Bandura, 1999; Eagly, 1987). We tested this possibility by assessing participants' proscriptive belief about the gender appropriateness of the novel occupations (e.g., "Who should do the job of being bensters?"). Results showed that among adolescents and adults, individuals endorsed more egalitarian attitudes toward jobs depicted with male workers than the identical jobs depicted with female workers. That is, jobs that are dominated by females are perceived as "only for females" and jobs that are dominated by males are perceived as "for both males and females." This discrepancy may be the result of individuals' belief that familiar feminine occupations are lower in status (e.g., money and power) than masculine occupations. This finding also may reflect society's acceptance of women in male-dominated professions (and masculine roles more generally), and society's reluctance to accept men in female-dominated professions (and feminine roles more generally).

The analyses most central to our hypotheses concerned the effects of sex of worker and values depicted in novel occupations on participants' occupational interests. With respect to sex of worker, across ages, participants showed significantly more interest in jobs depicted with workers of their own sex than in identical jobs depicted with workers of the other sex. These data are powerful because they indicate that sex of worker per se influences occupational interests, even when all other job characteristics are identical.

With respect to values, experimentally manipulating the values afforded by novel occupations significantly affected individuals' job interests. In addition, as hypothesized, males and females responded differently to the value manipulation. Across all three age groups, males were significantly more interested in jobs that were depicted as high in money than they were in the identical jobs depicted as high in power, family, or altruism. Overall, females, too, were significantly more interested in jobs that were depicted as high in money than in identical jobs depicted as high in power or family; unlike men, however, females were not significantly more interested in high-paying jobs than they were in altruistic jobs. Furthermore, men (but not women) were more interested in jobs depicted as high in power and family than they were in jobs depicted as high in altruism. In sum, the affordance of altruism was associated with sex differences in occupational interest, producing high interest among females but not among males.

This finding is particularly powerful because it is the first to demonstrate that perceived affordance of occupational values per se (i.e., apart from links to the sex-type and duties associated with an occupation) directly affects individuals' interests in occupations. Furthermore, the causal effect of values differed among males and females, as expected. The affordance of altruism was linked to decreased interest among males but not among females. Interestingly, the effects of the manipulation of values were evident in all age groups included in this research Thus, the results suggest that the sex-type of occupations and values afforded by occupations play important roles in shaping children's, adolescents', and adults' occupational interests.

General Discussion

Gender segregation of occupations is pervasive in the United States and elsewhere. Researchers in developmental psychology have long posited that gender cognitions (e.g., gender stereotypes) play a role in shaping occupational interests. In contrast, researchers in vocational psychology have focused on occupational values as shaping occupational goals. In this article, we used both correlational and experimental designs to examine the roles of both factors.

Gender-related cognitions appear to play an important, causal role in individuals' occupational interests. Consistent with research in the gender development literature (e.g., Bigler & Liben, 1990; Liben & Bigler, 2002; Ruble et al., 2006), results from Study 1 indicated that, in general, males show greater interest in traditionally masculine than feminine occupations, whereas females show greater interest in traditionally feminine than masculine occupations. In addition, results from Study 2 indicated that cognitions about gender segregation of occupations (i.e., sex of workers) alone can play a direct role in shaping individuals' job interests. Specifically, participants at all ages were significantly more interested in jobs depicted with workers of their own sex than in identical jobs depicted with workers of the other sex. These data suggest that merely observing that an occupation is performed exclusively by workers of the other sex may well reduce individuals' likelihood of developing or expressing an interest in that field and suggest the existence of a self-perpetuating cycle that makes it difficult to significantly increase the number of gender-nontraditional workers within fields.

Occupational values also appear to play a role in individuals' occupational interests. Evidence from the current studies is consistent with earlier work (Eccles, 1994; Marini et al., 1996; Schulenberg et al., 1993) in suggesting that sex differences in occupational values may lead to sex-typed occupational interests among adults. Furthermore, this pathway seems to develop gradually, beginning in childhood and strengthening across adolescence. Data from Study 1, for example, indicated that sex differences in occupational values (with males tending to endorse money and power values, and females tending to endorse altruistic and family values) were evident in adults and, to a limited extent, in adolescents, but not children. Furthermore, values differentially predicted individuals' interest in familiar masculine and feminine occupations (Study 1) and, perhaps most notably, predicted males' reported interest in entirely new occupations (Study 2). The finding that males—but not females—expressed greater interest in occupations that were depicted as affording money than those depicted as affording altruism is interesting and may help to explain the types of gender segregation typical of the workforce. The finding is consistent, for example, with research indicating that females are often more interested than males in cross-sex-typed jobs as youth and more willing to work in such jobs in adulthood (Lavine, 1982; Leung & Harmon, 1990; Liben & Bigler, 2002).

Although children did not show gender-differentiated values, their values predicted their interests differently as a function of gender. Males—at all ages—showed stronger interest in jobs depicted as affording high pay than identical jobs depicted as afford helping others. Why and how might boys come to value money more strongly than altruism? It seems possible that children are knowledgeable about the higher average earnings of male than female workers and internalize the value of money accordingly. It also seems possible that males and females might, over the course of childhood and adolescence, come to value different types of altruistic activities. Therefore, for example, although both sexes may value helping, females may prefer to provide emotional or instructional assistance (e.g., as in social work, psychology, and teaching), whereas males may prefer to provide instrumental assistance (as in fire fighting, surgery, or law). Emotional or instructional assistance may be more salient when responding to questions about "helping," and may be less strongly associated with money, than instrumental assistance.

It is also important to acknowledge limitations of this work. The cross-sectional design provides only a start toward understanding the pathways between early gender cognitions and values and later occupational choices. Longitudinal designs have increasingly been used with adolescent and college samples (Watt & Eccles, 2008), and should be extended to younger samples. Furthermore, we examined only four values (i.e., money, power, family, altruism) here. Future research should examine the role that other values have in shaping individuals' occupational interests, perhaps especially those values likely to be salient to children (e.g., adventure, danger, fame). Another limitation concerns our participants' work histories. Participants were not members of the full-time workforce, and thus they lacked experiences that may lead individuals to think more deeply about, and prioritize more strongly, their occupational values. Finally, it should be acknowledged that different patterns between age groups could, in part, reflect the different racial demographics of the three groups. Although the samples, in general, were diverse, they were not perfectly matched across ages and studies. Work indicates that race, ethnicity, and socioeconomic status—like gender—are also important for understanding occupational aspirations (Bigler, Averhart, & Liben, 2003; Cook, Church, Ajanaku, Shadish, & Kim, 1996; for a review, see Hughes & Bigler, 2008) and thus the role of these factors should also be explored in future research on developing values. Thus, although the research reported here provides some insights into the factors that affect individuals' occupational interests, it is also clear that additional research will be needed before the gender-differentiated world of work will be understood and, perhaps, thereby more readily open to educational interventions.

References

Bigler, R. S., Averhart, C. J., & Liben, L. S. (2003). Race and the workforce: Occupational status, aspirations, and stereotyping among African American children. *Developmental Psychology*, *39*, 572–580.

Bigler, R. S., & Liben, L. S. (1990). The role of attitudes and interventions in gender-schematic processing. *Child Development, 61*, 1440–1452.

Blakemore, J. E. O., Berenbaum, S. A., & Liben, L. S. (2009). *Gender development*. New York: Psychology Press.

Brown, D. (2002). The role of work and cultural values in occupational choice, satisfaction, and success: A theoretical statement. *Journal of Counseling & Development, 80*, 48–56.

Bussey, K., & Bandura, A. (1999). Social cognitive theory of gender development and differentiation. *Psychological Review, 106*, 676–713.

Cinamon, R. G., & Rich, Y. (2002). Gender differences in the importance of work and family roles: Implications for work-family conflict. *Sex Roles, 47*, 531–541.

Cook, T. D., Church, M. B., Ajanaku, S., Shadish, W. R., & Kim, J. R. (1996). The development of occupational aspirations and expectations among inner-city boys. *Child Development, 67*, 3368–3385.

Crouter, A. C., Bumpus, M. F., Maguire, M. C., & McHale, S. M. (1999). Linking parents' work pressure and adolescents' well being: Insights into dynamics in dual earner families. *Developmental Psychology, 35*, 1453–1461.

Eagly, A. H. (1987). *Sex differences in social behavior: A social-role interpretation*. Hillsdale, NJ: Erlbaum.

Eccles, J. S. (1987). Gender roles and women's achievement-related decisions. *Psychology of Women Quarterly, 11*, 135–171.

Eccles, J. S. (1994). Understanding women's educational and occupational choices: Applying the Eccles et al. model of achievement-related choices. *Psychology of Women Quarterly, 18*, 585–609.

Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., et al. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives* (pp. 75–146). San Francisco: W. H. Freemen.

Erb, T. O. (1983). Career preferences of early adolescents: Age and sex differences. *Journal of Early Adolescence, 3*, 349–359.

Gregg, C. H., & Dobson, K. (1980). Occupational sex role stereotyping and occupational interests in children. *Elementary School Guidance & Counseling, 15*, 66–75.

Hartung, P. J., Porfeli, E. J., & Vondracek, F. W. (2008). Career adaptability in childhood. *Career Development Quarterly, 57*, 63–74.

Heilman, M. (1979). High school students' occupational interests as a function of projected sex ratios in male-dominated occupations. *Journal of Applied Psychology, 64*, 275–279.

Herzog, A. R. (1982). High school seniors' occupational plans and values: Trends in sex differences 1976 through 1980. *Sociology of Education, 55*, 1–13.

Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments* (3rd ed.). Odessa, FL: Psychological Assessment Resources.

Hughes, J. M., & Bigler, R. S. (2008). The impact of race on children's occupational aspirations. In S. M. Quintana & C. McKown (Eds.), *The handbook of race, racism, and the developing child* (pp. 397–423). Hoboken, NJ: Wiley.

Ji, P. Y., Lapan, R. T., & Tate, K. (2004). Vocational interests and career efficacy expectations in relation to occupational sex-typing beliefs for eighth grade students. *Journal of Career Development, 31*, 143–154.

Johnson, M. K. (2001). Job values in the young adult transition: Stability and change with age. *Social Psychology Quarterly, 64*, 297–317.

Judge, T. A., & Bretz, R. D. (1992). Effects of work values on job choice decisions. *Journal of Applied Psychology, 77*, 261–271.

Knoop, R. (1991). Work values and job satisfaction. *Journal of Psychology, 128*, 683–690.

Konrad, A. M., Ritchie, J. E., Lieb, P., & Corrigall, E. (2000). Sex differences and similarities in job attribute preferences: A meta-analysis. *Psychological Bulletin, 126*, 593–641.

Lavine, L. (1982). Parental power as a potential influence on girls' career choice. *Child Development, 53*, 658–663.

Lee, C. C. (1984). Work values of rural Black, White, and Native American adolescents: Implications for contemporary rural school counselors. *Counseling & Values, 28*, 63–71.

Leung, S. A., & Harmon, L. W. (1990). Individual and sex differences in the zone of acceptable alternatives. *Journal of Counseling Psychology, 37*, 153–159.

Liben, L. S., & Bigler, R. S. (2002). The developmental course of gender differentiation: Conceptualizing, measuring, and evaluating constructs and pathways. *Monographs of the Society for Research in Child Development, 67*(2, Serial No. 269).

Liben, L. S., Bigler, R. S., & Krogh, H. R. (2001). Pink and blue collar jobs: Children's judgments of job status and job aspirations in relation to sex of worker. *Journal of Experimental Child Psychology, 79*, 346–363.

Lindsay, P., & Knox, W. E. (1984). Continuity and change in work values among young adults: A longitudinal study. *American Journal of Sociology, 89*, 918–931.

Marini, M. M., Fan, P., Finley, E., & Beutel, A. M. (1996). Gender and job values. *Sociology of Education, 69*, 49–65.

McMahon, M., & Watson, M. (2008). Introduction to the special section: Children's career development: Status quo and future directions. *Career Development Quarterly, 57*, 4–6.

Plost, M., & Rosen, M. J. (1974). Effect of sex of career model on occupational preference of adolescents. *AV Communication Review, 22*, 41–50.

Post-Kammer, P. (1987). Intrinsic and extrinsic work values and career maturity of 9th- and 11th-grade boys and girls. *Journal of Counseling and Development, 65*, 420–423.

Rowe, R., & Snizek, W. (1995). Gender differences in work values: Perpetuating the myth. *Work and Occupations, 22*, 215–229.

Ruble, D. N., Martin, C. L., & Berenbaum, S. A. (2006). Gender development. In W. Damon & R. M. Lerner (Series Eds.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (6th ed., pp. 858–932). Hoboken, NJ: Wiley.

Ryckman, R. M., & Houston, D. M. (2003). Value priorities in American and British female and male university students. *Journal of Social Psychology, 143*, 127–138.

Schulenberg, J., Goldstein, A. E., & Vondracek, F. W. (1991). Gender differences in adolescents' career interests: Beyond main effects. *Journal of Research on Adolescence, 1*, 37–61.

Schulenberg, J., Vondracek, F. W., & Kim, J. (1993). Career certainty and short-term changes in work values during adolescence. *Career Development Quarterly, 41*, 268–284.

Simpkins, S. D., Davis-Kean, P., & Eccles, J. S. (2006). Math and science motivation: A longitudinal explanation of the links between choices and beliefs. *Developmental Psychology, 42*, 70–83.

Singer, S. L., & Stefflre, B. (1954). The relationship of job values and desires to vocational aspirations of adolescents. *Journal of Applied Psychology, 38*, 419–422.

Summers, L. H. (2005). Remarks at NBER Conference on Diversifying the Science & Engineering Workforce. Retrieved August 14, 2007, from http://www.president.harvard.edu/speeches/2005/nber.html

Thomas, V. G. (1986). Career aspirations, parental support, and work values among Black female adolescents. *Journal of Multicultural Counseling and Development, 14*, 177–185.

Thomas, C., Benzeval, M., & Stansfeld, S. (2007). Psychological distress after employment transitions: The role of subjective financial position as a mediator. *Journal of Epidemiology & Community Health, 61*, 48–52.

Touhey, J. C. (1974). Effects of additional men on prestige and desirability of occupations typically performed by women. *Journal of Applied Social Psychology, 4*, 330–335.

Tracey, T. G. J., & Ward, C. C. (1998). The structure of children's interests and competence perceptions. *Journal of Counseling Psychology, 45*, 290–303.

U.S. Department of Commerce, Bureau of the Census. (2005). *2000 Census of Population*. Washington, DC: U.S. Government Printing Office.

Vida, M., & Eccles, J. (2003, April). Predicting mathematics-related career aspirations and choices. Paper presented at the meeting of the Society for Research in Child Development, Tampa, FL.

Watt, H. G., & Eccles, J. S. (2008). *Gender and occupational outcomes: Longitudinal assessment of individual, social, and cultural influences*. Washington, DC: American Psychological Association.

Weisgram, E. S., & Bigler, R. S. (2006). Girls and science careers: The role of altruistic values and attitudes about specific tasks. *Journal of Applied Developmental Psychology, 27*, 326–348.

Zytowski, D. G. (1970). The concept of work values. *Vocational Guidance Quarterly, 18*, 176–186.

Appendix

Occupational Values Scale for Children

All questions were structured as follows: "Some people have jobs that let them ______; other people have jobs that do not let them ______. How much would you like a job that lets you ______?" For each of 16 items, the following job qualities were inserted into the stem question: 1) let them make the rules for other people; 2) let them help other people; 3) let them quit for a while when they have kids and then they can come back to work when the kids are big; 4) let them earn enough money to buy a large house; 5) let them make important decisions; 6) let them take a day or two off of work to spend time with their family; 7) let them earn a great deal of money; 8) let them help people who need it, like people who are sick, poor, and homeless; 9) let them go to work *and* have lots of time with their family; 10) let them be in charge; 11) help others feel better; 12) let them earn more and more money each year; 13) let them be helpful to others; 14) let them be the boss; 15) let them work part-time (only sometimes, like half-days) when they have young kids; 16) let them take home a lot of money each week.

Child Development, May/June 2010, Volume 81, Number 3, Pages 797–810

The Head Bone's Connected to the Neck Bone: When Do Toddlers Represent Their Own Body Topography?

Celia A. Brownell, Sara R. Nichols,
and Margarita Svetlova
University of Pittsburgh

Stephanie Zerwas
University of North Carolina–Chapel Hill

Geetha Ramani
University of Maryland

Developments in very young children's topographic representations of their own bodies were examined. Sixty-one 20- and 30-month-old children were administered tasks that indexed the ability to locate specific body parts on oneself and knowledge of how one's body parts are spatially organized, as well as body-size knowledge and self-awareness. Age differences in performance emerged for every task. Body-part localization and body spatial configuration knowledge were associated; however, body topography knowledge was not associated with body-size knowledge. Both were related to traditional measures of self-awareness, mediated by their common associations with age. It is concluded that children possess an explicit, if rudimentary, topographic representation of their own body's shape, structure, and size by 30 months of age.

Children begin learning about their own bodies as newborns. Within a few hours of birth, neonates can tell if the hand caressing their cheek belongs to someone else or is their own (Rochat & Hespos, 1995). As infants use their bodies to engage the world—moving through space, watching their own hands and feet, playing with objects and people—they discover how their bodies move, what their bodies are capable of, and how their bodies and body parts relate to other things in the world (see Adolph & Berger, 2006, for a review). Thus, infants differentiate their bodies and actions from the physical and social world very early, developing a prereflective, ''tactile, auditory, and kinesthetic . . . bodily self'' (Butterworth, 1995, p. 93) over the course of the 1st year (Rochat, 1995).

This implicit, perceptually specified bodily self becomes explicit and available to conscious awareness beginning in the 2nd year of life as toddlers become able to take their own bodies and actions as objects of reflective thought (Bertenthal & Rose, 1995; Moore, 2007; Piaget, 1952; Rochat, 2001). Recent studies have shown that children become consciously aware of the size of their own bodies and of their bodies as potential obstacles or impediments late in the 2nd year of life, with development continuing into at least the 3rd year (Brownell, Zerwas, & Ramani, 2007; Moore, Mealiea, Garon, & Povinelli, 2007). The purpose of the current research was to build on and extend these findings to study very young children's explicit awareness of their body's shape and spatial configuration, sometimes termed *body topography* (e.g., Reed, 2002). To do so, we adapted tasks used to index such knowledge in adults and administered them to children between 20 and 30 months of age. This is the age when body self-awareness has been previously studied and when other aspects of objective self-awareness as traditionally measured are also developing, including mirror self-recognition, self-reference, and self-conscious emotions (e.g., Lewis & Ramsay, 2004).

This research was supported in part by grants from the National Institute of Child Health and Human Development, HD043971 and HD055283, to C.A.B. Portions of this research were presented at meetings of the Society for Research in Child Development and the International Conference on Infant Studies. We thank the children and parents who volunteered their time to participate in this research. Special thanks also to Rita Balaporia, Kacie Bartosavage, Prema Chaudhari, Sabrina DellaMea, Kristen Frey, Charla Newhouse, Yuliya Rinberg, Lauren Shuck, Alina Volper, and Diana Whalen for assistance with data collection and coding, to Alexis Brownell for help with artwork and stimulus preparation, and to Stephanie Sivers for help with manuscript preparation.

Correspondence concerning this article should be addressed to Celia A. Brownell, Department of Psychology, University of Pittsburgh, Pittsburgh, PA 15260. Electronic mail may be sent to brownell@pitt.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8103-0010

Dimensions of Body Representation

The conceptual framework underlying the current work rests on the proposal that children's earliest explicit representations of their own bodies are constructed along several distinct dimensions (Brownell et al., 2007). Two dimensions are the focus of the current study: (a) the body as a physical entity with objective attributes such as height and width (body-size awareness) and (b) the body as an organized collection of body parts with particular, stable spatial relations among them (body topography). This framework derives from empirical and theoretical work in cognitive neuroscience, adult neuropsychology, and developmental psychology.

Cognitive neuroscience researchers have confirmed that an abstract spatial representation of one's own body is associated with particular regions in the intact adult brain (Arzy, Thut, Mohr, Michele, & Blanke, 2006; Chaminade, Meltzoff, & Decety, 2005; Churchland, 2002; Devue et al., 2007; Ehrsson, Kito, Sadato, Passingham, & Naito, 2005; LeClec et al., 2000). Importantly, visual representations of the body in the brain are dissociable from movement-related representations of the body (e.g., Sugiura et al., 2006; Urgesi, Candidi, Ionta, & Aglioti, 2007) and also from the cortical network involved in the human mirror neuron system (see Peelen & Downing, 2007, for a review). Thus, the form of the human body is represented independently of the body's actions in the adult brain, raising the possibility that they are developmentally distinct as well.

Converging evidence from neuropsychological research with brain-injury patients has also shown that there are several distinct systems dedicated to the representation of one's own body (e.g., Schwoebel & Coslett, 2005; Sirigu, Grafman, Bressler, & Sunderland, 1991). For example, patients with damage to the left parietal area often cannot point to their own body parts (autotopagnosia) when an examiner names them (e.g., *elbow*, *chin*), even though they can locate body parts on animals, name human body parts after hearing them defined, and identify individual parts of inanimate objects such as bicycles (Guariglia, Piccardi, Puglisi, & Traballesi, 2002; Semenza, 1988). Such findings suggest a disruption in patients' topographic representation of their own body, which is dissociable from semantic representations of body parts. Other patients with left parietal damage find it especially difficult to imitate meaningless gestures (ideomotor apraxia) when the gesture is positioned at specific locations on the body. For example, a patient might watch an examiner model a gesture next to the ear only to imitate it next to the nose (Buxbaum & Coslett, 2001; Goldenberg, 1996). Because meaningless gestures directed to one's own body must be represented in terms of particular spatial configurations of the body and its parts, such errors suggest disrupted knowledge of the structure of the human body (Chaminade et al., 2005; Goldenberg, 1997; Goldenberg & Karnath, 2006). Notably, these sorts of impairments are not routinely associated with deficits in general motor function or the ability to perform everyday motor activities such as getting dressed or moving through space, again consistent with a distinction between visuospatial and movement-related representations of one's body. That these impairments occur in otherwise normally self-aware individuals further suggests that body self-awareness is a distinct aspect of objective self-awareness. However, research from adult neuroscience and neuropsychology cannot tell us how body self-awareness develops or how it might connect with other elements of the developing self.

Development of Body Self-Awareness

The world is represented from both first-person and third-person perspectives, and these two perspectives appear to become integrated relatively late in infancy, in the latter part of the 2nd year (Barresi & Moore, 1996; James, 1890; Kagan, 1981). This developmental achievement may be especially relevant for children's early body representations. To reflect consciously on one's own body shape and configuration requires integrating one's subjective, first-person experience of oneself with an objective, third-person perspective (e.g., Keysers & Gazzola, 2007; Moore, 2007; Povinelli & Cant, 1995).

Body topography. Slaughter and her colleagues (Heron & Slaughter, 2008; Slaughter & Heron, 2004) have systematically examined infants' developing third-person perspective on the human body as revealed in their knowledge about the body's canonical configuration. In a series of studies, they have established that 15- to 18-month-old infants can visually discriminate noncanonical body configurations (e.g., arms protruding from head instead of torso) from typical body shapes, but in a structured object-examination procedure infants do not reliably categorize typical versus scrambled bodies until 24 months. When an explicit categorization judgment was required, only 30-month-old girls were able to categorize typical versus scrambled bodies correctly; boys could categorize typical and scrambled cars by 24 months but could not categorize human body configurations at either 24 or 30 months (Heron & Slaughter, 2008). Thus, a detailed visuospatial representation of the human body becomes available in the 2nd year of life, when infants recognize the body as having a particular configuration, shape, and structure and are sensitive to the canonical spatial relations among body parts. During the 3rd year of life, knowledge of human body configuration becomes more explicit. However, there is no empirical work addressed to children's corresponding spatial representations of their *own* bodies; the primary aim of the current study was to fill this gap.

To this end, we examined the development of 1- and 2-year-olds' topographic knowledge of their own bodies. Two components corresponding to those studied with adults were the focus: the locations of one's own body parts (body-part localization) and the spatial relation among one's own body parts (body configuration knowledge). Because own body topographic representations have not been studied previously in young children, it is also important to know whether these two putative components cohere, that is, whether children develop a higher order, more general topographic self-map that includes both the specific locations of their body parts and the spatial relations among them. It was therefore a further aim to determine whether body-part localization and body configuration knowledge are related.

Body size. Building on research showing that toddlers sometimes try to fit their bodies into miniature objects that are much too small for them (DeLoache, Uttal, & Rosengren, 2004) or try to fit through a much too narrow opening (Garon & Moore, 2002), we recently examined children's body-size errors between 17 and 30 months of age across multiple tasks (Brownell et al., 2007). Although most of the children made such errors, frequency of errors declined with age as children became more likely to take their own size into account when trying to fit into very small doll clothes, trying to squeeze through a small door to reach their mothers on the other side, or playing with doll-size toys or furniture. A third aim of the current research was to replicate those results, given the novelty of the tasks and the potential importance of the findings.

The fourth aim of the study was to evaluate the potential correspondence between body-size awareness and body topography knowledge. If explicit awareness of one's own body size, shape, and structure is, indeed, emerging in this period, it is important to determine whether and how these early developments are interrelated. Finally, given that scholars have suggested that reflective self-awareness may ground body self-awareness (e.g., Moore, 2007), or vice versa (Povinelli & Cant, 1995), the final aim of the study was to examine associations between dimensions of body self-awareness and more traditional measures of reflective self-awareness.

We expected to find age differences for each dimension of body self-awareness over the 2nd year of life, partly based on prior empirical findings, and partly because this is the period when reflective self-awareness first appears and undergoes significant developmental change (Lewis & Brooks-Gunn, 1979; Nielsen, Dissanayake, & Kashima, 2003). We further expected that measures of body-part localization and body configuration would cohere as components of the larger construct of body topography. However, we had no strongly grounded expectations for potential associations between body topography knowledge and body-size awareness. It is possible that these are somewhat distinct dimensions of developing body self-awareness as suggested by the fact that in adults, distortions in the body image that center on body size (e.g., some eating disorders, body dysmorphic disorder) do not covary with disruptions in spatial awareness of one's body (e.g., autotopagnosia, hemi-neglect). Finally, we expected that both body-size awareness and body topography knowledge would be associated with standard measures of reflective self-awareness insofar as body self-awareness would seem to be grounded in objectification of the self more generally.

Method

Participants

Participants were 61 children at two ages: 20 months ($M = 20.4$ months, $SD = 0.99$; 14 girls, 16 boys) and 30 months ($M = 29.6$ months, $SD = 1.2$; 13 girls, 18 boys). Families were recruited from a medium-sized urban area and nearby suburbs, and were predominantly middle class and Caucasian (86% Caucasian, 7% Asian, 5% African American, 2% Hispanic). All children were walking and were healthy and developing normally by parent report. Three additional children participated but their data could not be used because of motor delay, language delay, or probable autism spectrum disorder.

General Procedure

Procedures took place in two adjacent playrooms and were videotaped for later coding. Parents remained with their children at all times. A primary experimenter (E) administered all procedures with an assistant experimenter (AE) for some tasks. Children became comfortable with both experimenters during a warm-up free play period at the beginning of the session.

Five tasks were administered to assess body self-awareness. Two tasks evaluated children's knowledge of their own body topography, and three tasks evaluated children's understanding of their body size. Each task required 5–10 min to administer; all children received all tasks. The standard mirror self-recognition (rouge) task was also administered midway through the session, and parents completed questionnaires to assess children's self-understanding.

Body Topography Tasks

Two new tasks were developed to provide nonverbal assessments of young children's ability to represent the locations and spatial arrangement of their body parts. These were adapted from tasks used to study body representation deficits in adults with focal brain damage, as reviewed earlier. One task requires adults to point to their own body parts as the examiner names them. Another task requires adults to imitate meaningless gestures directed to particular locations on the body. Toddler versions of both tasks were created and extensively pilot-tested with 40 children between 18 and 32 months of age to provide the assessments used in the current study.

Sticker task: Body-part localization. This task assessed children's ability to locate particular body parts on themselves. Children were asked to place a sticker on an *unnamed* body location on themselves after watching E place a sticker at that location on AE. No body-part names or labels were used, to avoid confounds with developing lexical knowledge.

The child and both experimenters were seated on the floor, with AE directly across from the child, approximately 50 cm away. E drew the child's attention to AE and said, ''Watch, I'm going to put a sticker right there on [name],'' as she placed the sticker on AE. She then said again, ''See, I put a sticker right there,'' while she pointed to or touched the location on AE. She then handed a sticker to the child and said, ''Now you put your sticker on you right there, so it's just like [name]. You put your sticker right there on you.''

The first location was always the nose since previous research has shown that it is one of the first body parts learned: Eighty-five percent of 18-month-olds and 100% of 24-month-olds can point to their nose on request (Witt, Cermak, & Coster, 1990). Based on existing research on body-part naming (MacWhinney, Cermak, & Fisher, 1987; Witt et al., 1990) and parent report from pilot data, the six most often known body locations among 18- to 24-month-olds were determined (nose, hand, foot, head, back, and neck) as well as the six least known locations (forehead, wrist, elbow, calf, temple, and nape). We included both to guard against either floor or ceiling effects; we neither expected every child to know the most familiar, nor did we expect every child to fail the least familiar. These constituted the 12 standard locations presented to each child. Each location was demonstrated one at a time by E. The sticker remained fully visible on AE until the children placed their own sticker on themselves.

Children's sticker placements were scored from video records for accuracy (within 1 in. of location; adjacent location on same body part, different body part, nonbody location, refuse) by two trained coders. Because young children's motor limitations sometimes made it difficult for them to place the stickers with high precision and because it is unlikely that children of this age understand the notion of spot-on accuracy, accurate and adjacent locations were combined to yield the total number of ''correct'' sticker locations. Proportion correct was the measure used in analyses (number of correct sticker placements divided by 12). Interrater agreement was calculated for each coder with a master coder (one of the authors) on 31% of the sample (11% and 20% respectively; $\kappa s = 1.0$ and .96).

Meaningless gesture imitation: Body-part Configuration. This task assessed children's ability to represent the spatial relation among their own body parts. In adults, being able to maintain a meaningless hand gesture and to position it on the body requires explicit awareness and representation of one's individual body parts and the spatial relations among them (Goldenberg, 1996; Goldenberg & Karnath, 2006). Children were asked to imitate a meaningless gesture, a closed fist, positioned at particular body locations. Pilot testing had determined that children within this age range could imitate a closed fist without difficulty.

The child was seated at a child-sized table (65 cm × 45 cm × 40 cm). E sat diagonally to the child, approximately 40 cm to the child's right. She explained to the child that they would play a turn-taking game and then demonstrated one of nine different actions. Each action was demonstrated twice while the child watched. The child was encouraged to imitate after the second demonstration and was given approximately 30 s to do so. If the child did not imitate, the action was demonstrated once again with another opportunity to imitate. If the child failed to imitate again, the next action was demonstrated. For six of the actions, the experimenter demonstratively placed her closed fist on a specific body location (forearm, stomach, cheek, chin, forehead, or top of head). To control for individual differences in motor control, imitative ability, and/or motivation, the child was also asked to imitate three different actions with objects, all directed to the table top (slide a wooden block on the table, pat the table with a plastic dishwashing scrubby, and pound the table with a wooden cylinder). Each of these actions required the child to place the hand approximately in a fist shape to grasp the object, and all of the objects were nonspecific to the action demonstrated, thus all were novel action–object combinations.

Children's performance was scored from video records by two trained coders for the number of correct imitations. A response was considered correct if the child imitated the experimenter's action at the same location (within 1 in., and without penalty for confusions between left and right), using the same hand configuration (for gestures) and orientation (for gesture and objects) as the experimenter's. Interrater agreement for the number of correct trials was calculated on 21% of the sample ($\kappa = .91$).

Body-Size Tasks

Three tasks indexed children's ability to reason about their own body size relative to objects in the world. These were identical to those used in our prior study (Brownell et al., 2007) since one aim of the current study was to replicate those findings. Additional details about the tasks can be found in that publication.

Doll clothes. In this task, children were offered much too small doll clothes to wear. The child watched as the experimenter dressed a 30-cm doll in a doll-sized hat, jacket, and shoes. After placing each piece of clothing on the doll, the experimenter handed the child an identical piece of doll clothing and said simply, "Here's your [hat, jacket, shoe]." Attempts by children to put the doll clothes on themselves as if the clothing were full-sized were scored as errors (coding details given later).

Door choice. This task required the child to choose one of two doors to reach a parent on the other side of a 1 m × 2 m foamboard wall, only one of which the child could fit through (adapted from Garon & Moore, 2002). One door was a 30 cm × 30 cm hole in the wall, with a 10-cm square window above it through which the child could see the parent. Children could easily crawl through this door. The second door was tall and narrow (10 cm wide × 80 cm tall), and the child could also see the parent through this hole; however, it was much too narrow for the child's body. The child was placed equidistant between the doors, and the parent then called the child. Once the child had joined the parent on the other side, the experimenter called the child to return. Children's attempts to squeeze their bodies through the too small door were scored as errors.

Replica toys. In this task (adapted from DeLoache et al., 2004), children played with a standard set of child-sized toys (Little Tikes™ slide, toddler chair, cozy coupe car; Little Tikes Company, Hudson, OH) and then several minutes later played with a separate set of small, doll-sized replicas of the same toys; these were approximately 8–14 cm high, and one fifth to one sixth the height of the child-sized toys. Children were free to play with the replica toys however they wished for approximately 5 min. Attempts to use the replica toys as if they were full-sized, without considering that their bodies are too big for such small toys (e.g., trying to sit in the doll chair), were scored as errors.

Body-size awareness measures. Children's behavior was coded from the video records by two trained coders. An error was defined as a serious, nonpretend attempt, based on the child's action, facial expression, and effort. For example, a serious attempt to wear the doll jacket meant attempting with clear effort to put the hand and forearm fully into the sleeve; a serious attempt to sit in the doll-sized replica chair involved positioning the body appropriately and then sitting with one's full weight on the chair with a neutral facial expression (and perhaps surprise at the outcome). Each independent attempt was counted as an error. Interobserver reliability for the number of errors on each task was calculated on 21% of the sample; because the codes were not mutually exclusive and exhaustive, percent agreements rather than Kappas were calculated for the individual tasks, and ranged from 82% to 89%.

Reflective Self-Awareness

Three independent, age-appropriate measures of reflective self-awareness were obtained using established procedures: (a) mirror self-recognition, (b) parental report of the child's self-description and self-evaluation, and (c) parental report of the child's comprehension and production of personal pronouns. Mirror self-recognition was assessed using the standard, well-replicated ''rouge task'' (Amsterdam, 1972) in which the parent surreptitiously placed a dab of red lipstick on the child's nose during free play. A brief period of continuing play ascertained that the child neither noticed the lipstick nor tried to remove it spontaneously. The experimenter then held a 30 cm × 30 cm mirror in front of the child so that a full view of his or her own face was visible, drawing the child's attention to the mirror. If the child did not respond after approximately 10 s, the experimenter asked ''Who's there?'' while pointing to the mirror; if the child did not attend after an additional 10 s the experimenter asked ''What's on his or her nose?'' again pointing to the mirror.

Children's mirror-related behavior was independently coded from video records by two trained coders. Behaviors included touching or turning away from the mirror, saying own name, touching own nose, touching other parts of the face, attempting or requesting to remove the lipstick, and self-conscious behavior such as embarrassment. Children's performance was scored on a 3-point scale (after Asendorpf, Warkentin, & Baudonniere, 1996; Courage, Edison, & Howe, 2004; Rogdon & Kurdek, 1977): *pass* (2) if they said their own name, touched their own nose, or attempted or requested to remove the lipstick (48% of 20-month-olds, 81% of 30-month-olds); *ambiguous* (1) if they touched other parts of their face but not their nose, touched their nose upon hearing the word *nose* in the experimenter's attention-eliciting cue, or were self-conscious in front of the mirror (24% of 20-month-olds, 19% of 30-month-olds); *fail* (0) if they engaged in any other behavior, including saying ''baby,'' touching the mirror, ignoring it, or turning away from it without self-conscious behavior (28% of 20-month-olds, 0% of 30-month-olds). Interobserver reliability was calculated on 21% of the sample ($\kappa = .88$).

Parents completed the UCLA Self-Understanding Questionnaire developed for this age group (Stipek, Gralinski, & Kopp, 1990), with 17 items rated on a 3-point scale (0 = *definitely not*, 1 = *sometimes/just starting to*, 2 = *definitely*). Items tap self-recognition, self-description, and self-evaluation (e.g., recognizes self in photos; uses own name; says ''me,'' ''mine''; uses terms like *good* or *bad* about self); Cronbach's alpha = .91. Parents also completed the toddler form of the MacArthur Communicative Development Inventory (MCDI; Fenson et al., 1994). Because we were interested in words referring to self and other, and because use of personal pronouns has previously been shown to relate to mirror self-recognition in this age group (Lewis & Ramsay, 2004), we used the *Pronouns* subscale of the MCDI for which the parent indicates whether the child understands (a) or understands and says (b) eight personal pronouns (e.g., *her, you, me*); Cronbach's alpha = .93. The three measures of reflective self-awareness were significantly intercorrelated (*r*s = .35–.59, *p*s = .001–.005), so were standardized and summed to produce a single, global measure of reflective self-awareness (Cronbach's alpha = .77).

Results

Preliminary analyses indicated that there were no sex differences on any measure; thus, analyses were collapsed over sex. Table 1 shows mean values and standard errors for all variables. Substantive analyses were conducted to address several questions. The first set of analyses examined age differences for each aspect of body knowledge (body-part localization, body configuration, body size) and for reflective self-awareness. The second set tested associations between body-part localization and body configuration to determine whether there was evidence for a general topographic representation of one's own body in children of this age. The third set examined associations between body-size awareness and body topographic knowledge to determine whether they are part of a more general body-awareness construct or whether they are distinct. The final set of analyses examined associations between elements of body knowledge and reflective self-awareness. Not all children completed all tasks; thus, *n*s vary slightly across analyses.

Age Differences

Univariate or multivariate analyses of variance (ANOVAs) with two levels of age (20 and

Table 1
Mean Values and Standard Errors for Body Knowledge and Self-Awareness Measures With Significance Test Results for Age Differences (ANOVA)

	All children ($n = 61$)	20-month-olds ($n = 30$)	30-month-olds ($n = 31$)	F	p value
Body-size tasks (no. of errors)					
Doll clothes	2.12 (0.31)	2.93 (0.36)	1.37 (0.22)	6.58	.013
Door choice	1.63 (0.23)	2.32 (0.26)	0.93 (0.15)	10.05	.003
Replica toys	2.88 (0.42)	4.11 (0.46)	1.73 (0.33)	8.49	.005
Body-part localization (proportion correct)					
Sticker task	0.27 (0.28)	0.19 (0.20)	0.35 (0.33)	4.87	.031
Body-part configuration (proportion correct)					
Imitation: Gestures to self	0.17 (0.05)	0.08 (0.04)	0.26 (0.04)	10.03	.003
Imitation: Object actions to table	0.77 (0.03)	0.76 (0.07)	0.78 (0.07)	0.004	*ns*
Reflective self-awareness (z score)[a]	−0.04 (0.32)	−1.83 (0.22)	1.69 (0.23)	60.89	.001

[a] MSR + UCLA self-understanding + CDI personal pronouns.

30 months) were conducted as appropriate on the measures of each dimension of body representation and on the composite measure of reflective self-awareness. As expected, we found systematic age differences for each measure.

Body-part localization. Children were asked to place stickers on 12 different body locations on themselves after watching a sticker placed at each location on another person. A univariate ANOVA with proportion of correct sticker placements as the dependent variable yielded a significant age effect, $F(1, 59) = 4.87$, $p = .031$, $\eta^2 = .08$. Older children placed nearly twice as many stickers on or adjacent to the correct body part as did younger children (see Table 1), although it should also be noted that even older children were still far from perfect. Some children frequently placed stickers on surrounding objects instead of on themselves or refused to place several stickers anywhere ($n = 15$). When these children were removed from the analysis, average performance rose (20 months, $M = 0.23$; 30 months, $M = 0.47$), and the significant age effect remained, $F(1, 44) = 9.57$, $p = .003$, $\eta^2 = .18$.

Body configuration. Children were asked to imitate a meaningless gesture directed to their own bodies at six different locations; in the comparison condition they were asked to imitate three different actions with an object directed to a table top. A repeated measures ANOVA with location (self, table) as the within-subjects factor and age as the between-subjects factor was conducted on the proportion of correct imitations (see Table 1). A significant effect for location emerged, $F(1, 55) = 17.20$, $p < .001$, $\eta^2 = .76$, as well as a marginally significant interaction between age and location, $F(1, 55) = 3.69$, $p = .06$, $\eta^2 = .06$. As evident in Table 1, younger children performed as well as older children when imitating actions directed toward the table but performed less accurately than older children when they had to imitate a gesture directed to their own body.

Body size. The number of times children attempted to fit themselves into small doll clothes, through small openings, or onto or into miniature toys that were much too small was recorded. A multivariate ANOVA with the three measures of body-size awareness (doll clothes errors, door choice errors, and replica toys errors) as dependent variables yielded a significant age effect, multivariate $F(3, 52) = 6.48$, $p = .001$, $\eta^2 = .27$. The follow-up univariate tests for age differences on each dependent variable were also significant (see Table 1, for Fs and p values). Thus, on all three measures, 30-month-olds produced significantly fewer errors than did 20-month-olds, replicating the findings of Brownell et al. (2007). Also as in Brownell et al., children's performance in the current study was correlated across the three tasks (rs = .22–.35, ps = .10–.006).

Reflective self-awareness. A composite measure of reflective self-awareness was created from mirror self-recognition scores, children's self-description and self-evaluation, and their personal pronoun comprehension and use. The ANOVA yielded a significant age effect, $F(1, 59) = 60.89$, $p < .001$, $\eta^2 = .51$, as expected (see Table 1).

Evidence for a Topographic Body Representation

To determine whether children's body-part localization and body configuration knowledge

Table 2
Intercorrelations Among Body Knowledge and Self-Awareness Measures

	Personal pronouns	UCLA self-understanding	Mirror self-recognition	Reflective self-awareness[a]	Sticker task[b]	Imitation: Gestures to self[b]	Imitation: Object actions to table[b]
UCLA self-understanding	.81***						
Mirror self-recognition	.35**	.42***					
Reflective self-awareness[a]	.80***	.82***	.41**				
Sticker task[b]	.29*	.30*	.21†	.33**			
Imitation: Gestures to self[b]	.28*	.40**	.26*	.36**	.53***		
Imitation: Object actions to table[b]	.03	.14	.31*	.21†	.35**	.38**	
Body-size tasks (sum of errors)	−.21†	−.37**	−.28*	−.33*	−.13	−.13	−.08

Note. $df = 57$ to 61, depending on the number of children completing a given task.
[a]Summary score. [b]Proportion correct.
†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

constituted related components of a higher order topographic representation of their own bodies, Pearson correlations were calculated between children's performance on the sticker task and their performance on the imitation task (see Table 2 for correlation coefficients). The significant association suggests that young children do possess a general topographic representation of their own bodies.

Unexpectedly, however, children's ability to locate stickers on their own body was also associated with their imitation of object actions directed to the table (see Table 2), suggesting that some factor related to general imitative skill or motivation might also be at work and could explain the above association. To control for this possibility, we recalculated the correlation between children's score for imitating gestures to their own bodies and their score for placing stickers on particular body parts, partialing out their ability to imitate actions with objects; the correlation remained significant, partial $r(57) = .43$, $p = .001$. Finally, we further partialed age; children's ability to imitate meaningless gestures to their own body remained significantly associated with their ability to locate particular body parts on themselves above and beyond both general imitative ability and age, partial $r(56) = .32$, $p = .014$.

Associations Between Topographic Body Representations and Body Size Awareness

To determine to what extent topographic body knowledge and body-size awareness are distinct or related, Pearson correlations were calculated between a composite measure of body-size awareness (the total number of body-size errors across all tasks; Cronbach's alpha = .61) and the measures of each of the other two components of body awareness, that is, body-part localization and body configuration. Neither of these correlations was significant (see Table 2 for correlation coefficients), including when age was partialed from the calculations. A composite score for body topography knowledge was also created as the sum of z scores for the sticker and gestural imitation tasks, and was likewise unrelated to the number of body-size errors. It thus appears that there are two distinct aspects of early body self-awareness, one related to children's knowledge of their own body size and one related to the topographic representation of their body and its parts.

Associations Between Body Representation and Reflective Self-Awareness

Finally, to determine whether the different aspects of body self-awareness indexed in the current study were associated with reflective self-awareness as traditionally assessed, Pearson correlations were calculated between the composite self-awareness measure and the composite measures for body-size awareness (number of errors), $r(59) = -.33$, $p = .01$, and body topography knowledge (number correct), $r(64) = .38$, $p = .002$. The same results were obtained when the self-awareness composite was correlated with the separate measures for body-part localization and body configuration knowledge individually (see Table 2), as well as when the composite measure of reflective self-awareness was decomposed into its individual constituents. When age was partialled, associations became nonsignificant ($rs < .20$). Thus, as expected, reflective self-awareness was associated with each dimension of body self-awareness studied here;

however, these relations were carried by their common association with age.

Discussion

Early developments in children's explicit, conscious representations of their own body's objective physical characteristics such as shape, size, and configuration were explored in the current study. Development of body self-awareness is relatively uncharted despite continuing interest in the growth of self- and other-understanding (Brownell & Kopp, 2007), and the potential for identifying the roots of childhood body image in infancy (Smolak, 2004). We report evidence that children's topographic representation of their own bodies emerges between 20 and 30 months of age.

We further found that this aspect of early body knowledge appears to be constituted of two related dimensions. One dimension reflects knowledge about where specific body parts are located on one's own body even when these are not named or previously known. The second reflects knowledge about how one's body is spatially configured, that is, how one's body parts are arranged or spatially organized in relation to one another. Both appear to be still relatively immature at 2½ years of age, pointing to continuing development over the preschool years. We also found that children become aware of their own body size during the 2nd year of life, replicating previous findings (Brownell et al., 2007). Interestingly, children's awareness of their body's topographic characteristics was unrelated to their body-size awareness. Both were related, however, to a multifaceted measure of self-recognition and self-awareness, albeit mediated by their common associations with age.

Together with other recent research (e.g., Brownell et al., 2007; Moore et al., 2007; Nielsen, Suddendorf, & Slaughter, 2006), the findings from the current study confirm that body self-awareness is a unique aspect of the development of an objective, conscious self. The results further suggest that the child's developing body image, even in its initial and relatively primitive form, is constituted of multiple, distinct dimensions. In particular, previous research showed that body self-awareness emerges in the 2nd year as explicit awareness of one's physical characteristics like size, mass, or solidity, and of being able to enter into object–object relations with other physical objects by serving as an obstacle, tool, encumbrance, or container (Brownell et al., 2007; Moore et al., 2007). The current results suggest that it is somewhat later, in the 3rd year of life, that awareness of the topographic or spatial organization of one's body also appears. By age 2½, then, children have a beginning knowledge of their own body's size, shape, and structure.

Of course, toddlers' conscious awareness of their own body size and configuration does not arise de novo, but emerges out of and builds on an extensive implicit and perceptually based bodily self established over infancy. This includes sensitivity to body dimension parameters (e.g., Adolph & Avolio, 2000) and to one's own body parts such as hands (McCarty, Clifton, & Collard, 1999). For example, before 12 months of age infants can scale motor actions like reaching or throwing to take account of object size, position, orientation, speed, or function in motor planning tasks (e.g., Claxton, Keen, & McCarty, 2003; Lockman, Ashmead, & Bushnell, 1984; von Hofsten, Vishton, Spelke, Feng, & Rosander, 1998; Witherington, 2005). With the advent of reflective self-awareness in the latter half of the 2nd year, this implicit, sensorimotor knowledge of one's body becomes available to conscious awareness. As Moore (2007) has argued, it is during this period that infants' developing first-person knowledge of their own bodies becomes integrated with their developing third-person knowledge of others' bodies. This, in turn, permits conscious reflection on body characteristics of both self and others, thereby also grounding the subsequent development of the body image.

Own Body Topography

To study toddlers' own body representations, we created age-appropriate modifications of tasks used by investigators of adults' representations of their own body configuration (e.g., Goldenberg & Karnath, 2006; Guariglia et al., 2002; Schwoebel, Buxbaum, & Coslett, 2004). In one task, children were asked to locate body parts by placing a sticker on a particular location on their own body after watching an adult place a sticker at that location on an assistant's body without naming the location or body part. On average, 20-month-old children were able to locate only two or three common body parts (e.g., nose, hand, foot), whereas 30-month-olds were able to locate four or five (e.g., nose, hand, foot, head, neck). These findings indicate that children's knowledge of the specific locations of their own body parts is quite primitive at 20 months of age, but is well in evidence by 30 months, even though still immature.

A similar procedure was used by DeLoache and Marzolf (1995) to study preschoolers' ability to use a doll to represent their own body. In that study, when an experimenter placed stickers on four specific places on the child's body (forearm, knee, foot, and neck) and then asked the child to place stickers in the same places on the doll, 32-month-old children placed only 1.6 stickers correctly out of the 4 possible. Moreover, children found it equally difficult to perform the procedure from doll to self. This suggests that for 2-year-olds, mapping between one's own body and a doll's body may be even more difficult than mapping between one's own body and an adult's body, possibly because of additional cognitive demands in forming a representational relation between self and a pretend object. More relevant for the current study, these findings confirm that during the 3rd year of life children's spatial knowledge of their own bodies undergoes substantial progress.

In contrast to the findings from these two studies, lexical acquisition studies indicate that 12-month-olds can identify two body parts on average when asked to point to them by name (Witt et al., 1990) and that by 18–24 months of age children can name and point to several specific body parts on themselves (Bretherton, McNew, & Beeghly-Smith, 1981; Kopp, 1994). However, learning to point to specific body parts on command may be a result of highly practiced routines and thus need not depend on or reflect a more general representation of the spatial configuration of the body, including where various unrehearsed body parts are located and how body parts are arranged and organized in relation to one another. This is why, in the current study, we included body parts for which children did not yet have names (e.g., elbow) and refrained from naming those body parts for which they were likely to possess a label (e.g., nose, foot). We are confident, therefore, that children's performance in the current study was not a result of stereotyped routines for pointing to specific, well-rehearsed body locations on request. Instead, we conclude that children's explicit self-representations undergo significant developmental change at the end of the 2nd year of life and into the 3rd year.

In the second task used to index their body topography knowledge, children were asked to imitate a meaningless gesture (closed fist) placed at different locations on the body (e.g., top of head, stomach). This requires consideration of multiple relation among specific body parts, both those involved in making and maintaining the gesture itself, and the location on the body at which the gesture is placed (Goldenberg, 1996). We found that 30-month-old children were able to position their fists accurately on their own bodies more than 3 times as often as were younger children. Importantly, younger and older children were equivalent in their ability to imitate object-related actions directed to a table top. Thus, it was not general imitative ability that limited the younger children's performance, but the specific ability to imitate a gesture directed to locations on their own bodies. Like the results for children's knowledge of where their body parts are specifically located, these findings indicate that knowledge of how their own body is spatially organized is also quite rudimentary in 1-year-olds, becoming evident in young 2-year-olds, though clearly still primitive even then.

These data are consistent with previous findings that imitation of unfamiliar and/or nonvisible bodily movements and gestures is more difficult than imitation of familiar movements or imitation with objects, both in normal children under 24 months and in children with autism (Abravanel, Levan-Goldschmidt, & Stevenson, 1976; Brownell, 1988; Jones, 2007; Masur, 1993; Masur & Ritz, 1984; Rogdon & Kurdek, 1977; Rogers, Hepburn, Stackhouse, & Wehner, 2003; Stone, Ousley, & Littleford, 1997). They are also consistent with several recent anecdotal reports of children's inaccurate matching responses when they imitate behavior with or on specific body parts. For example, when 14- or 18-month-old infants watched an adult illuminate a light by using her forehead or elbow to touch the light, they imitated body-part-specific actions only approximately; that is, they used their mouth, chin, cheek, or even ear to touch the light, rather than the forehead as modeled (Herold & Akhtar, 2008; Southgate, Gergely, & Csibra, 2009), or they used the back of the hand instead of the elbow as modeled (Herold & Akhtar, 2008). In a different study, when 12- and 18-month-old infants were asked to imitate actions with objects on their own body after watching an experimenter model the behavior on herself (e.g., roll a car on an arm, tap a plastic ring against the cheek), children often failed to use the correct body part when they imitated; for example, they would tap the ring on their chin instead of their cheek (Carpenter, Tomasello, & Striano, 2005). Together, these data converge on the conclusion that the ability to represent explicitly the spatial relation among one's own body parts begins to take form late in the 2nd year of life, between 20 and 30 months of age.

Despite the rather pronounced demand differences in the sticker (body-part localization) and imitation (body configuration) tasks, children who performed better on one task also performed better on the other, even with both age and general imitative ability controlled. Performance on body-part localization and meaningless gesture imitation tasks is also strongly associated in adults with brain damage ($r = .78$; Schwoebel et al., 2004), and seems to call on similar regions of the brain (Goldenberg & Karnath, 2006). Our data thus suggest that by 30 months of age young children, like adults, begin to access consciously a general representation of their own bodies' spatial organization. Even these 2-year-olds, however, were far from ceiling on either task, which indicates that development of this complex aspect of body self-awareness must be ongoing well past the toddler years.

Own Body-Size Awareness

The current study provided an independent replication of our previous work in which we found developmental changes over the 2nd year in children's knowledge of their body size (Brownell et al., 2007). Interestingly, a recent study has shown that young children between 16 and 60 months of age are poor at scaling their reaching actions to the size of their own hands; 1- and 2-year-olds, in particular, tried repeatedly to fit their hands through impossibly small openings to reach something on the other side, even after their hands became stuck (Ishak & Adolph, 2008). Thus, the ability to imagine and attend to one's own body size begins to emerge in the 2nd year of life and appears to continue developing over the preschool years.

The fact that body-size awareness and body topography knowledge were not associated with each other suggests that these body representations are somewhat distinct and may develop partly independently of one another. Christie and Slaughter (2009) similarly found no associations between infants' sensorimotor body representations and their visuospatial body representations, prompting those authors to suggest that different body representation systems may initially be differentiated, then become integrated over the course of infancy and early childhood. Another possibility is that there are additional task demands, or perhaps affordances, unique to some of the body-size awareness tasks used here and by Brownell et al. (2007). If fitting through doorway openings, for example, relies on children's affordance detection rather than (or in addition to) explicit knowledge of their own body size, then failure may reflect other performance factors such as immature impulse control, or other demand factors such as low cost for failure. Teasing apart potential contributions to failure on such tasks holds promise for better understanding the mechanisms that contribute to eventual success, and the circumstances in which explicit body knowledge can or must be invoked.

Reflective Self-Awareness and Body Self-Awareness

Body-size awareness and body topography knowledge were associated with traditional measures of reflective self-awareness, including mirror self-recognition, self-description and self-evaluation, and language about self and other. However, this association was carried by the strong common associations with children's age. These findings contrast with those of Moore et al. (2007) who found that toddlers' ability to move themselves off a rug attached to the rear of a shopping cart before pushing it was related to mirror self-recognition, even with age controlled. However, the association with self-recognition was modest in that study, leading the authors to propose that body self-awareness and mirror self-recognition tap different, but related, aspects of developing self-awareness. We draw the same conclusion from the findings of the current study, noting further that mirror self-recognition is mastered by most children by 24–30 months of age, whereas children's body self-awareness remains rather rudimentary at 30 months of age, with development extending into early childhood. Reflective self-awareness may thus be a necessary foundation, but not sufficient to account specifically for growth and change in body-size awareness or topographic body knowledge.

This proposal is consistent with recent arguments for a somewhat extended developmental course for other complex aspects of self-recognition (Suddendorf, Simcock, & Nielsen, 2007). For example, there is a developmental lag of about 1 year between children's mirror self-recognition and their ability to recognize themselves in real-time video, even after controlling for differences in the perceptual features of the two types of display. Suddendorf et al. (2007) speculate that this developmental asynchrony is explained by an immature mental self-image among younger children, which is lacking in the abstraction necessary to support a general, intermodal representation of their own appearance. It would be interesting in future research to determine whether the child's primitive body image, as reflected in the kinds of tasks used in the current study, consolidates during the same period in development as video and shadow self-recognition, which are presumably also based on an objective, integrated, spatially configured own body representation. Such a general representation of one's appearance may serve as a developmental link between the perceptually specified bodily self of infancy and the later emergence of the body image in early childhood.

References

Abravanel, E., Levan-Goldschmidt, E., & Stevenson, M. B. (1976). Action imitation: The early phase of infancy. *Child Development*, *47*, 1032–1044.

Adolph, K., & Avolio, A. (2000). Walking infants adapt locomotion to changing body dimensions. *Journal of Experimental Psychology: Human Perception & Performance*, *26*, 1148–1166.

Adolph, K., & Berger, S. (2006). Motor development. In W. Damon & R. Lerner (Series Eds.) & D. Kuhn & R. S. Siegler (Vol. Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (6th ed., pp. 161–213). New York: Wiley.

Amsterdam, B. (1972). Mirror self-image reactions before age two. *Developmental Psychobiology*, *5*, 297–305.

Arzy, S., Thut, G., Mohr, C., Michele, C., & Blanke, O. (2006). Neural basis of embodiment: Distinct contributions of temporoparietal junction and extrastriate body area. *Journal of Neuroscience*, *26*, 8074–8081.

Asendorpf, J. B., Warkentin, V., & Baudonniere, P. M. (1996). Self-awareness and other-awareness. II: Mirror self-recognition, social contingency awareness, and synchronic imitation. *Developmental Psychology*, *32*, 313–321.

Barresi, J., & Moore, C. (1996). Intentional relations and social understanding. *Behavioral & Brain Sciences*, *19*, 107–122.

Bertenthal, B., & Rose, J. (1995). Two modes of perceiving the self. In P. Rochat (Ed.), *The self in infancy: Theory and research* (pp. 303–324). Amsterdam: Elsevier.

Bretherton, I., McNew, S., & Beeghly-Smith, M. (1981). Early person knowledge as expressed in gestural and verbal communication. In M. Lamb & L. Sherrod (Eds.), *Infant social cognition: Empirical and theoretical considerations* (pp. 333–353). Hillsdale, NJ: Erlbaum.

Brownell, C. (1988). Combinatorial skills: Converging developments over the second year. *Child Development*, *59*, 675–685.

Brownell, C., & Kopp, C. (2007). Transitions in toddler socioemotional development: Behavior, understanding, relationships. In C. Brownell & C. Kopp (Eds.), *Socioemotional development in the toddler years: Transitions and transformations* (pp. 1–42). New York: Guilford.

Brownell, C., Zerwas, S., & Ramani, G. (2007). "So big": The development of body self-awareness in toddlers. *Child Development*, *78*, 1426–1440.

Butterworth, G. (1995). An ecological perspective on the origins of self. In J. L. Bermudez, A. J. Marcel, & N. Eilan (Eds.), *The body and the self* (pp. 87–106). Cambridge, MA: MIT Press.

Buxbaum, L. J., & Coslett, H. (2001). Specialized structural descriptions for human body parts: Evidence from autotopagnosia. *Cognitive Neuropsychology*, *18*, 289–306.

Carpenter, M., Tomasello, M., & Striano, T. (2005). Role reversal imitation and language in typically-developing infants and children with autism. *Infancy*, *8*, 253–278.

Chaminade, T., Meltzoff, A., & Decety, J. (2005). An fMRI study of imitation: Action representation and body schema. *Neuropsychologia*, *43*, 115–127.

Christie, T., & Slaughter, V. (2009). Exploring links between sensori-motor and visuo-spatial body representations in infancy. *Developmental Neuropsychology*, *34*, 448–460.

Churchland, P. (2002). Self-representation in nervous systems. *Science*, *296*, 308–310.

Claxton, L., Keen, R., & McCarty, M. (2003). Evidence of motor planning in infant reaching behavior. *Psychological Science*, *14*, 354–356.

Courage, M. L., Edison, S. C., & Howe, M. L. (2004). Variability in the early development of visual self-recognition. *Infant Behavior and Development*, *27*, 509–532.

DeLoache, J., & Marzolf, D. (1995). The use of dolls to interview young children: Issues of symbolic representation. *Journal of Experimental Child Psychology*, *60*, 155–173.

DeLoache, J., Uttal, D., & Rosengren, K. (2004). Scale errors offer evidence for a perception-action dissociation early in life. *Science*, *304*, 1027–1029.

Devue, C., Collette, F., Balteau, E., Degueldre, C., Luxen, A., Maquet, P., et al. (2007). Here I am: Cortical correlates of visual self-recognition. *Brain Research*, *1143*, 169–182.

Ehrsson, J., Kito, T., Sadato, N., Passingham, R., & Naito, E. (2005). Neural substrate of body size: Illusory feeling of shrinking of the waist. *PloS Biology*, *3*(12), e439.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D., Pethick, S. J., et al. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *59*(5, Serial No. 242).

Garon, S., & Moore, C. (2002, April). *Development of the objective self*. Paper presented at the International Conference on Infant Studies, Toronto, ON.

Goldenberg, G. (1996). Defective imitation of gestures in patients with damage in the left or right hemispheres. *Journal of Neurology, Neurosurgery, and Psychiatry*, *61*, 176–180.

Goldenberg, G. (1997). Disorders of body perception. In T. Feinberg & M. Farah (Eds.), *Behavioral neurology and neuropsychology* (pp. 289–296). New York: McGraw-Hill.

Goldenberg, G., & Karnath, H. (2006). The neural basis of imitation is body part specific. *Journal of Neuroscience*, *26*, 6282–6287.

Guariglia, C., Piccardi, M., Puglisi, A., & Traballesi, M. (2002). Is autotopoagnosia real? EC says yes. A case study. *Neuropsychologia, 40*, 1744–1749.

Herold, K., & Akhtar, N. (2008). Learning from a third-party interaction: Relations with self-recognition and perspective taking. *Journal of Experimental Child Psychology, 101*, 114–123.

Heron, M., & Slaughter, V. (2008). Toddlers' categorization of typical and scrambled dolls and cars. *Infant Behavior & Development, 31*, 374–395.

Ishak, S., & Adolph, K. E. (2008). *Fitting through openings: Body scaling in infants, children, and adults*. Manuscript in preparation.

James, W. (1890). *Principles of psychology*. New York: Holt.

Jones, S. (2007). Imitation in infancy: Development of mimicry. *Psychological Science, 18*, 593–598.

Kagan, J. (1981). *The second year: Emergence of self-awareness*. Cambridge, MA: Harvard University Press.

Keysers, C., & Gazzola, V. (2007). Integrating simulation and theory of mind: From self to social-cognition. *Trends in Cognitive Science, 11*, 194–196.

Kopp, C. (1994). *Baby steps: Your child's behavior in the first two years*. New York: Freeman.

LeClec, G., Dehaene, S., Cohen, L., Mehler, J., Dupoux, E., Poline, J., et al. (2000). Distinct cortical areas for names of numbers and body parts independent of language and input modality. *NeuroImage, 12*, 381–391.

Lewis, M., & Brooks-Gunn, J. (1979). *Social cognition and the acquisition of self*. New York: Plenum.

Lewis, M., & Ramsay, D. (2004). Development of self-recognition, personal pronoun use, and pretend play during the second year. *Child Development, 75*, 1821–1831.

Lockman, J., Ashmead, D., & Bushnell, E. (1984). The development of anticipatory hand orientation during infancy. *Journal of Experimental Child Psychology, 37*, 176–186.

MacWhinney, K., Cermak, S., & Fisher, A. (1987). Body part identification in 1- to 4-year-old children. *American Journal of Occupational Therapy, 41*, 454–459.

Masur, E. F. (1993). Transitions in representational ability: Infants' verbal, vocal, and action imitation during the second year. *Merrill-Palmer Quarterly, 39*, 437–456.

Masur, E., & Ritz, E. (1984). Patterns of gestural, vocal and verbal imitation in infancy. *Merrill-Palmer Quarterly, 30*, 369–392.

McCarty, M., Clifton, R., & Collard, R. (1999). Problem solving in infancy: Emergence of an action plan. *Developmental Psychology, 35*, 1091–1101.

Moore, C. (2007). Understanding self and others in the second year. In C. Brownell & C. Kopp (Eds.), *Transitions in early socioemotional development: The toddler years* (pp. 43–65). New York: Guilford.

Moore, C., Mealiea, J., Garon, N., & Povinelli, D. (2007). The development of body self-awareness. *Infancy, 11*, 157–174.

Nielsen, M., Dissanayake, C., & Kashima, Y. (2003). A longitudinal investigation of self-other discrimination and the emergence of mirror self-recognition. *Infant Behavior & Development, 26*, 213–226.

Nielsen, M., Suddendorf, T., & Slaughter, V. (2006). Mirror self-recognition beyond the face. *Child Development, 77*, 176–185.

Peelen, M., & Downing, P. (2007). The neural basis of visual body perception. *Nature Reviews: Neuroscience, 8*, 636–648.

Piaget, J. (1952). *The origins of intelligence in children*. New York: International Universities Press.

Povinelli, D., & Cant, J. (1995). Arboreal clambering and the evolution of self-conception. *Quarterly Review of Biology, 70*, 393–421.

Reed, C. (2002). What is the body schema? In A. Meltzoff & W. Prinz (Eds.), *The imitative mind: Development, evolution, and brain bases* (pp. 233–245). Cambridge, UK: Cambridge University Press.

Rochat, P. (Ed.). (1995). Early objectification of the self. In *The self in infancy: Theory and research* (pp. 53–71). Amsterdam: Elsevier.

Rochat, P. (2001). Origins of self-concept. In G. Bremner & A. Fogel (Eds.), *Blackwell handbook of infant development* (pp. 191–212). Malden, MA: Blackwell.

Rochat, P., & Hespos, S. (1995). Differential rooting response by neonates: Evidence for an early sense of self. *Early Development and Parenting, 6*, 150–158.

Rogdon, M., & Kurdek, L. (1977). Vocal and gestural imitation in 8- 14- and 20-month-old children. *Journal of Genetic Psychology, 131*, 115–123.

Rogers, S. J., Hepburn, S. L., Stackhouse, T., & Wehner, E. (2003). Imitation performance in toddlers with autism and those with other developmental disorders. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 44*, 763–781.

Schwoebel, J., Buxbaum, L., & Coslett, H. (2004). Representations of the human body in the production and imitation of complex movements. *Cognitive Neuropsychology, 21*, 285–298.

Schwoebel, J., & Coslett, H. (2005). Evidence for multiple, distinct representations of the human body. *Journal of Cognitive Neurosicience, 17*, 543–553.

Semenza, C. (1988). Impairment in localization of body parts following brain damage. *Cortex, 24*, 443–449.

Sirigu, A., Grafman, J., Bressler, K., & Sunderland, T. (1991). Multiple representations contribute to body knowledge processing. *Brain, 114*, 629–642.

Slaughter, V., & Heron, M. (2004). Origins and early development of human body knowledge. *Monographs of the Society for Research in Child Development, 69*(2, Serial No. 276).

Smolak, L. (2004). Body image in children and adolescents: Where do we go from here? *Body Image, 1*, 15–28.

Southgate, V., Gergely, G., & Csibra, G. (2009). Does the mirror neuron system and its impairment explain human imitation and autism? In J. Pineda (Ed.), *Mirror neuron systems: The role of mirroring processes in social cognition* (pp. 331–354). New York: Humana Press.

Stipek, D. J., Gralinski, J. H., & Kopp, C. B. (1990). Self-concept development in the toddler years. *Developmental Psychology*, *26*, 972–977.

Stone, W. L., Ousley, O. Y., & Littleford, C. D. (1997). Motor imitation in young children with autism: What's the object? *Journal of Abnormal Child Psychology*, *25*, 475–485.

Suddendorf, T., Simcock, G., & Nielsen, T. (2007). Visual self-recognition in mirrors & live videos: Evidence for a developmental asynchrony. *Cognitive Development*, *22*, 185–196.

Sugiura, M., Sassa, Y., Jeong, H., Miura, N., Akitsuki, Y., Horie, K., et al. (2006). Multiple brain networks for visual self-recognition with different sensitivity for motion and body part. *NeuroImage*, *32*, 1905–1917.

Urgesi, C., Candidi, M., Ionta, S., & Aglioti, S. (2007). Representation of body identity and body actions in extrastriate body area and ventral premotor cortex. *Nature Neuroscience*, *10*, 30–31.

von Hofsten, C., Vishton, P., Spelke, E., Feng, Q., & Rosander, K. (1998). Predictive action in infancy: Tracking and reaching for moving objects. *Cognition*, *67*, 255–285.

Witherington, D. (2005). The development of prospective grasping control between 5 and 7 months. *Infancy*, *7*, 143–161.

Witt, A., Cermak, S., & Coster, W. (1990). Body part identification in 1- to 2-year old children. *American Journal of Occupational Therapy*, *44*, 147–153.

Child Development, May/June 2010, Volume 81, Number 3, Pages 811–825

I Like Me If You Like Me: On the Interpersonal Modulation and Regulation of Preadolescents' State Self-Esteem

Sander Thomaes, Albert Reijntjes, and Bram Orobio de Castro
Utrecht University

Brad J. Bushman
University of Michigan & VU University Amsterdam

Astrid Poorthuis
Utrecht University

Michael J. Telch
University of Texas at Austin

This experiment tested whether peer approval and disapproval experiences can cause immediate change in children's state self-esteem. Children's narcissistic traits and evaluator perceived popularity were examined as potential moderators. A total of 333 preadolescents (M = 10.8 years) completed personal profiles on the Internet that were ostensibly judged by a jury consisting of popular and unpopular peers. Participants randomly received negative, neutral, or positive feedback from the jury. Next, they could examine the feedback that each individual judge gave them. As expected, peer disapproval decreased self-esteem, especially in children high in narcissism. In contrast, peer approval increased self-esteem. Moreover, disapproved children's self-esteem recovery was dependent on the extent to which they subsequently viewed positive feedback from popular judges. These findings support sociometer theory.

> *As social beings we live with our eyes upon our reflection, but have no assurance of the tranquillity of the waters in which we see it.*
>
> —Charles Horton Cooley, 1902, p. 247

Is children's self-esteem dependent on how much they are liked by others? Classical theory holds that throughout development, children gradually come to internalize the opinions that others hold of them into more or less stable patterns of self-esteem (Cooley, 1902; Mead, 1934). Supporting this view, research shows that children's *trait* self-esteem (i.e., their enduring appraisal of value as a person) tends to reflect the cumulative appraisals of others in the past (e.g., Boivin, Hymel, & Hodges, 2001; Cole, Jacquez, & Maschman, 2001; Felson & Zielinski, 1989; Ladd & Troop-Gordon, 2003; but see Salmivalli & Isaacs, 2005, for an exception). Another classical view holds that self-esteem is open to temporary fluctuations, and may wax and wane from moment to moment around an "average tone of self-feeling" (James, 1890). This latter, fluctuating type of self-esteem is called *state* self-esteem. Surprisingly little is known about the extent to which children's state self-esteem is sensitive to appraisals from others in the here and now. One possibility is that children's state self-esteem is only marginally reactive to current approval experiences, because these experiences are embedded in a long history of previous social interactions. Another possibility is that children's state self-esteem is profoundly reactive to current approval experiences, because these experiences are highly salient and meaningful in the present. This study uses an in vivo, experimental peer evaluation design to examine children's interpersonal state self-esteem reactivity in real time. In addition, it examines some theoretically relevant factors that may influence children's state self-esteem reactivity, namely narcissism and the perceived popularity of the person providing the appraisal.

Interpersonal Appraisals and the Modulation of State Self-Esteem

A pivotal contemporary account of interpersonal influences on self-esteem is found in sociometer theory (Leary & Baumeister, 2000; Leary & Downs, 1995). The basic assumption underlying the theory is that human beings possess a pervasive drive to

---

Correspondence concerning this article should be addressed to Sander Thomaes, Department of Psychology, Utrecht University, P.O. Box 80.140, 3584 CS Utrecht, Netherlands. Electronic mail may be sent to s.thomaes@uu.nl.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8103-0011

be valued and accepted by others, a drive that evolved because our ancestors were more likely to survive and reproduce when they belonged to social groups (e.g., Ainsworth, 1989; Baumeister & Leary, 1995; Leary & Baumeister, 2000). Sociometer theory posits that self-esteem has evolved to (a) monitor how much one is valued and accepted by others, (b) evoke positive or negative feelings that signal change in others' approval, and (c) provide an outlook for how one may be approved in future interactions so that one can try to prevent possible disapproval. Thus, according to sociometer theory, interpersonally based changes in self-esteem are not just accidental by-products of positive or negative social interaction; they are central in enabling people to gain, maintain, and restore approval from others.

These functions of self-esteem are thought to operate in the short run and in the long run. Trait self-esteem is proposed to function as a long-term gauge that monitors the amount of approval obtained from others, instills enduring negative mood (e.g., depression) when one is disapproved by others, and motivates people to prevent long-term exclusion. State self-esteem is proposed to function as a short-term gauge that continuously monitors one's social environment and signals changes in immediate approval obtained from others. Social cues that connote approval or disapproval are assumed to *modulate* state self-esteem and trigger positive or negative affect (e.g., feelings of pride or shame), signaling that one's "relational value" is bound to change. The pleasure or pain associated with state self-esteem fluctuations should motivate people to prevent exclusion in near future interactions. Research involving adult participants supports both the trait and state self-esteem postulates of sociometer theory. Trait self-esteem is a direct function of the extent to which adults generally feel approved by others (e.g., Leary, Cottrell, & Phillips, 2001; Leary, Tambor, Terdal, & Downs, 1995). Interpersonally based changes in trait self-esteem elicit longer term changes in negative mood and one's outlook for others' approval (e.g., Johnson, Meyer, Winett, & Small, 2000). State self-esteem, in contrast, fluctuates according to how adults feel approved by others *in the moment*, and these fluctuations bring about short-term changes in negative mood and alert individuals to possible exclusion in the near future (e.g., Leary & Downs, 1995; Leary, Haupt, Strausser, & Chokel, 1998; Oosterwegel, Field, Hart, & Anderson, 2001).

Empirical evidence for sociometer theory in children, however, is incomplete. That is, research in children has focused almost exclusively on long-term trait self-esteem and its interpersonal determinants (e.g., Boivin et al., 2001; Felson & Zielinski, 1989; Ladd & Troop-Gordon, 2003), and has largely neglected short-term state self-esteem and its interpersonal determinants. This seems unfortunate, because the psychological foundation for sensitivity to others' appraisals is laid early in life (Rubin, Bukowski, & Parker, 2006). From middle childhood to preadolescence, two cognitive advances allow for interpersonal influences on self-esteem (Harter, 1999, 2006). First, improved perspective taking skills help preadolescents to better appreciate the appraisals of others. Second, the newfound ability to form higher order concepts allows preadolescents to make global self-evaluations (e.g., "I am satisfied with myself as a person") rather than more concrete and domain-specific self-evaluations (e.g., "I am good at soccer") that are typical for younger children. Because self-esteem is relatively unstable in preadolescence (Harter, 1999, 2006), one might expect ongoing appraisals of others to have a particular strong impact at this age. Appraisals of peers should be especially important because the peer group exerts an increasingly salient socializing function in preadolescence (Harter, Stocker, & Robinson, 1996; Rubin et al., 2006; Rudolph, Caldwell, & Conley, 2005).

A few studies have examined how children *think* their self-esteem depends on current peer appraisals (Burwell & Shirk, 2006; Harter & Whitesell, 2003; Harter et al., 1996; Rudolph et al., 2005). To our knowledge, however, only one study examined the self-esteem fluctuations that children actually *experience* following current peer appraisals, and the evidence from this single study was inconsistent with sociometer theory (Nesdale & Lambert, 2007). Specifically, Nesdale and Lambert (2007) found that experimentally manipulated social rejection triggered negative affect (e.g., sadness, anxiety) but triggered no changes in children's self-esteem. Two limitations of Nesdale and Lambert's pioneering work should be noted, however. First, children were asked to role play a situation in which they were socially rejected; the social rejection was not played out in vivo. Although children can be responsive to role play (Rubin et al., 2006), their experience of pretended social encounters does not necessarily correspond well to their experience of actual social encounters (Robinson & Clore, 2002). Second, a measure of *trait* self-esteem was used (sample item: "I think I can do things as well as anyone else"), whereas sociometer theory predicts that current interpersonal appraisals influence children's *state* self-esteem (Leary, 1999; Leary et al., 1995). The first goal of the present experiment is to examine whether peer approval and disapproval modulate preadolescents' state self-esteem using an in vivo experimental peer evaluation manipulation.

Individual Differences in the Interpersonal Modulation of State Self-Esteem

According to sociometer theory, interpersonal self-esteem reactivity likely is a universal human trait. Of course, this is not to say there are no individual differences in the degree to which children's self-esteem is dependent on others' appraisals. Whereas some children's state self-esteem may be stable and relatively unaffected by current peer approval or disapproval experiences, other children's state self-esteem may be unstable and highly reactive to those experiences (Rudolph et al., 2005). The logic underlying sociometer theory implies that these individual differences should be a direct function of the importance children attach to being valued by others (see also Harter, 1999). Children who are greatly concerned with being valued by others should theoretically show pronounced interpersonal self-esteem reactivity.

An excessive need for external validation is a hallmark feature of the trait of narcissism. Whereas narcissism is a personality disorder in its extreme form (*DSM–IV*; American Psychiatric Association, 1994), most current psychological research (including our own) focuses on ''normal narcissism,'' operationalized as a personality trait on which people in the general population vary (Morf & Rhodewalt, 2001; Raskin & Terry, 1988). ''Normal narcissists'' believe they are superior to others, yet at the same time they worry obsessively about what others think of them. They crave to be respected and admired by others, and tend to respond with intense emotions when they receive negative interpersonal feedback (Baumeister & Vohs, 2001; Morf & Rhodewalt, 2001; Rhodewalt & Morf, 1998). Thus, narcissistic self-esteem is not only grandiose but is also markedly vulnerable to negative social experience (Morf & Rhodewalt, 2001). Recent research found that individual differences in narcissistic traits can be meaningfully measured in children at least 8 years old (Thomaes, Stegge, Bushman, Olthof, & Denissen, 2008) and that narcissistic children are emotionally reactive to negative social events (Thomaes, Bushman, Stegge, & Olthof, 2008). The second goal of this experiment is to examine narcissism as a potential moderator of the interpersonal modulation of state self-esteem. To establish discriminant validity, the moderating effect of narcissism is contrasted with the moderating effects of self-perceived social acceptance and peer-perceived social preference. These variables were chosen as controls because they are often considered indexes of social vulnerability (e.g., McElhaney, Antonishak, & Allen, 2008; Patterson, Kupersmidt, & Griesler, 1990) but do *not* involve an excessive need for external validation, which is presupposed to be a key determinant of individual differences in interpersonal self-esteem reactivity.

Interpersonal Appraisals and the Regulation of State Self-Esteem

Similar to other evolutionary-based monitoring mechanisms (e.g., hunger), self-esteem can reach certain critical low levels at which it is experienced as unpleasant, which then motivate behaviors intended to decrease the unpleasantness. Just as hunger feels bad and therefore motivates people to seek for food, lowered state self-esteem feels bad (i.e., triggers emotional distress) and therefore may motivate people to seek praise and positive feedback from others (Leary & Baumeister, 2000). Theoretically, the more positive feedback that disapproved children with lowered state self-esteem obtain, the more their self-esteem should recover. We therefore propose that interpersonal appraisals are involved not only in the modulation of state self-esteem but also in the subsequent *regulation* of state self-esteem.

One might be inclined to assume that when disapproved children with lowered state self-esteem manage to subsequently acquire positive social feedback, they should invariably show pronounced increases in state self-esteem. Such a position may be oversimplified, however. From a young age, children attach differential importance to the appraisals that others hold of them, typically based on salient characteristics of others that cue group membership (Crick & Dodge, 1994; McAninch, Manolis, Milich, & Harris, 1993; Mrug & Hoza, 2007). In elementary school, one important characteristic that children use to judge each other is perceived popularity (LaFontana & Cillessen, 2002). Children who are perceived as popular exert a strong influence on their peers. They are on the top of the social hierarchy and to a large extent they determine the social rules according to which some children are valued and accepted whereas others are not (Adler & Adler, 1998; Hawley, Little, & Pasupathi, 2002; LaFontana & Cillessen, 2002). Thus, if state self-esteem functions as a monitor of children's relational value, it should be particularly reactive to appraisals provided by perceived popular peers. The third goal of this experiment is to examine how peer appraisals influence the regulation of disapproved children's lowered state self-esteem. In particular, it will be examined whether positive appraisals from perceived popular peers have a stronger regulatory impact on state self-esteem than appraisals from unpopular peers.

Present Study

An experimental peer evaluation manipulation was used to examine the modulation and subsequent regulation of children's state self-esteem. Children's narcissistic traits and evaluator perceived popularity were examined as potential moderators of children's state self-esteem reactivity. Participants ranged in age from 8 to 12 years. This age group was chosen for four reasons. First, children this age have acquired the abilities to appreciate others' appraisals of them and to evaluate their global worth as a person (Harter, 1999, 2006). Second, children this age attach great importance to the opinions and appraisals of peers, much more so than younger children (Damon & Hart, 1988; Harter, 1999, 2006; Rosenberg, 1979). Third, given their heightened concern with who is popular and who is not, children this age may be especially susceptible to the appraisals of perceived popular peers (Adler & Adler, 1998; LaFontana & Cillessen, 2002). Fourth, individual differences in narcissism can be reliably measured in children this age (Thomaes, Stegge, et al., 2008), who typically have outgrown the unrealistically positive self-views that are normative in younger children (Marsh, Craven, & Debus, 1998).

Peer approval and disapproval were manipulated using the Survivor Game, which is an ostensible Internet contest shown to be highly effective at inducing credible peer approval and disapproval experiences in children (e.g., Reijntjes, Dekovic, Vermande, & Telch, 2007; Reijntjes, Stegge, Meerum Terwogt, Kamphuis, & Telch, 2006). Participants created a personal profile containing self-disclosing information as well as a photograph of themselves. By random assignment, they received negative social feedback (i.e., the lowest likability score), neutral social feedback (i.e., neither the lowest or the highest likability score), or positive social feedback (i.e., the highest likability score) on their profile from an Internet peer jury. In a subsequent self-esteem regulation period, participants could examine the feedback that each individual judge gave them. The individual judges provided either positive or negative feedback and were either perceived popular or unpopular. Thus, whereas the manipulated jury feedback was between subjects, the individual judges' feedback that participants subsequently viewed was within subjects. State self-esteem was assessed at baseline (T1), immediately after the manipulated jury feedback (T2), and immediately after viewing the individual judges' feedback (T3).

Based on sociometer theory, we predicted that from T1 to T2 children's state self-esteem would significantly decrease after disapproval, would not change after neutral feedback, and would significantly increase after approval. We predicted that children showing largest decreases in state self-esteem following disapproval would be high in narcissism (but would not differ from others in self-perceived social acceptance and peer-perceived social preference). Although narcissists may be more reactive to negative than to positive social events (Bushman & Baumeister, 1998; Rhodewalt & Morf, 1998), we also explored whether children showing largest increases in state self-esteem following approval were high in narcissism. Regarding the interpersonal regulation of state self-esteem, we predicted that positive feedback from perceived popular judges would have strong regulatory effects. Specifically, we predicted that disapproved children would show more state self-esteem recovery from T2 to T3 the more time they spent examining positive feedback from perceived popular judges (relative to positive feedback from perceived unpopular judges, or negative feedback). We also predicted that disapproved children would generally spend more time examining positive feedback from popular judges than other types of feedback.

Method

Participants

Participants were 333 preadolescents (49% boys) who ranged in age from 8 to 12 years ($M = 10.8$, $SD = 0.9$). They were recruited from 10 randomly selected public elementary schools serving lower (i.e., 4.2% below average Dutch income) to upper (i.e., 7.3% above average Dutch income) middle-class communities in the Netherlands. The regions where the schools are located are representative of the Netherlands in terms of socioeconomic status, level of urbanization, and cultural or ethnical diversity. Most participants were ethnic Dutch (88%); others mainly had mixed cultural or ethnical origins. All participants received informed parental consent (consent letters were sent home from schools, were signed and returned; consent rates ranged between schools from 68% to 95%; *M* consent rate = 83%). Participants received a small gift (e.g., mechanical pens, markers) in exchange for their voluntary participation.

Measures

A few weeks prior to the experiment (*M* = 2 weeks, range = 1–3 weeks), participants completed measures of narcissism, self-perceived social acceptance, and a sociometric nomination procedure in their classes. Trained female psychology students provided instructions and answered questions. During the experiment, participants completed a measure of state self-esteem.

Narcissism

Narcissism was measured using the Dutch version of the 10-item Childhood Narcissism Scale (Thomaes, Stegge, et al., 2008). This psychometrically sound scale assesses grandiose views of self and inflated feelings of superiority and entitlement in children. Sample items include: ''Without me, our class would be much less fun,'' ''Kids like me deserve something extra,'' and ''I often succeed in getting admiration.'' Items are rated along a 4-point scale ranging from 0 (*not at all true*) to 3 (*completely true*). Responses were summed, with higher scores indicating higher levels of narcissism (Cronbach's $\alpha = .79$).

Self-perceived social acceptance. Self-perceived social acceptance was measured using the Dutch version of the six-item social acceptance subscale of Harter's Self-Perception Profile for Children (1985; translated and validated by Veerman, Straathof, Treffers, Van den Bergh, & ten Brink, 1997). This scale assesses children's evaluation of how well they can get along with their peers. Children first choose which of two contrasting descriptors describes themselves best (e.g., ''Some kids would like to have a lot more friends, but other kids have as many friends as they want''). They then indicate whether the chosen descriptor is ''sort of true'' or ''very true'' for them. Items are rated along a 4-point scale ranging from 0 to 3. Responses were summed, with higher scores indicating higher levels of self-perceived social acceptance (Cronbach's $\alpha = .70$).

Peer-perceived social preference. Peer-perceived social preference was measured using sociometric nominations (Coie, Dodge, & Coppotelli, 1982). Children nominated the children who they ''like most'' and ''like least'' in their classes. Following conventional procedures (e.g., Cillessen & Mayeux, 2004), the number of nominations children received was standardized within each classroom. Then, the ''like least'' standardized score was substracted from the ''like most'' standardized score, and the resulting score was restandardized to create a measure of social preference.

State self-esteem. During the experiment, participants completed a six-item state self-esteem measure at baseline (T1; Cronbach's $\alpha = .68$), immediately after the manipulated jury feedback (T2; Cronbach's $\alpha = .80$), and immediately after viewing the individual judges' feedback (T3; Cronbach's $\alpha = .78$). The measure was modeled after standard measures of trait self-esteem (i.e., global self-worth subscale of Harter's Self-Perception Profile for Children, 1985; Rosenberg's Self-Esteem Scale, 1965), but assessed children's overall sense of worth *right now, at the present time*. Items reflected both a positive sense of self (e.g., ''I feel good about who I am right now'') and a negative sense of self (e.g., ''I am dissatisfied with myself right now''; see the Appendix for all items). Items were rated along a 5-point scale ranging from 0 (*not at all*) to 4 (*very much*). Responses were summed with higher scores indicating higher levels of state self-esteem. In a recent study involving 405 children and young adolescents, our measure of state self-esteem was found to be strongly associatiated with a one-item pictorial measure of state self-esteem ($r = .71$, $p < .0001$; Thomaes, Bushman, Orobio de Castro, Cohen, & Denissen, 2009). In addition, significant associations in the predicted direction were found between our measure of state self-esteem and trait measures of self-esteem (Harter, 1988), depression (Kovacs, 1981), and social anxiety (La Greca & Stone, 1993).

Experimental Procedure

Participants were individually tested in a quiet room at their school by a female research assistant. First, they completed the T1 state self-esteem measure. Next, they were told they would be competing in a popularity contest on the Internet called Survivor Game (a bogus game created for research purposes). Participants were told they would play the game with four opponents of the same sex and age from different schools, and that all players would be evaluated by a panel of 16 peer judges. Each judge would give them a ''likability score'' between 0 (*not at all likable*) and 100 (*extremely likable*), and would also provide concrete written feedback. Participants were then told that we were interested in how the game ''would make them feel'' and what they ''would think about'' the judges' feedback. They were also informed that they were free to cease participation at any time they wanted and still receive their gift (2 children stopped before the game began). Next, participants completed a personal profile, so that the peer judges could ostensibly learn more about them. Questions concerned participants' hobbies, future occupation, things they like and dislike about themselves, how well they get along with other children, academic performance, and self-ratings on several personal traits (e.g., sense of humor, agreeableness, intelligence, trustworthiness). Participants were told that their personal profile and their picture would be posted on the Survivor Game Web page to be evaluated by the peer judges.

To enhance credibility, participants then viewed the personal profiles of their opponents, taken from actual profiles provided by participants in a different study (consent for using the profiles was obtained; names were changed). After viewing the last profile, a message appeared on screen indicating that the computer would now tally the judges' scores for each player. After a short waiting period, the names of the players with the highest and the lowest scores appeared in capital letters on screen. Participants were randomly assigned to one of three conditions. In the approval condition, the participant received the highest score, whereas one randomly chosen opponent received the lowest score. In the disapproval condition, the participant received the lowest score, whereas one randomly chosen opponent received the highest score. In the control condition, two randomly chosen opponents received the highest and lowest scores. Because participants allegedly scored in between they did not see their own name.

Immediately after receiving the jury results, participants completed the T2 state self-esteem measure. Next, an announcement appeared on screen informing participants that they would have 5 min to look over the feedback comments that the judges wrote about them. An overview screen appeared containing pictures of all 16 judges (8 boys and 8 girls). By clicking on the picture of a targeted judge, a new screen opened containing concrete feedback provided by that judge. The judges were counterbalanced to be either perceived popular or unpopular, and to provide either positive or negative feedback. The peer status of the judges was manipulated through the two key determinants of children's perceived popularity: their physical appearance and number of friends (LaFontana & Cillessen, 2002). Popular judges were good looking and wrote they had ''many friends,'' whereas unpopular judges were not good looking and wrote they had ''few friends.'' In a pilot study, 10 children (M = 10.8 years) rated the judges' pictures (accompanied by a written statement about their number of friends) on a scale from 0 to 10, and considered the popular judges as better looking ($M = 7.6$, $SD = 0.6$ vs. $M = 3.9$, $SD = 1.4$) and as ''probably more popular'' ($M = 7.7$, $SD = 0.7$ vs. $M = 3.9$, $SD = 1.1$) than the unpopular judges, $t(14) = 6.93$, $p < .0001$, $d = 3.46$, and $t(14) = 8.24$, $p < .0001$, $d = 4.12$, respectively. The valence of the judges' feedback (positive or negative) was manipulated through the likability ratings and concrete feedback comments they provided. Positive judges provided favorable ratings (M likability score = 80 of 100) and wrote only positive comments (e.g., ''I would like to be friends with this person''). Negative judges provided unfavorable ratings (M likability score = 40 of 100) and wrote only negative comments (e.g., ''I would not like to be friends with this person''). Regardless of the experimental condition, all participants saw the same feedback from the 16 peer judges (i.e., 4 popular judges gave positive feedback, 4 popular judges gave negative feedback, 4 unpopular judges gave positive feedback, 4 unpopular judges gave negative feedback). The time (in seconds) that participants spent examining the four different types of feedback was recorded by the computer. Immediately after the feedback viewing period, participants completed the T3 state self-esteem measure.

Finally, participants were thoroughly debriefed. They were informed that the other players, the judges, and the evaluations they received were entirely fictitious. They were also informed about the study purposes, and the need for deception. They were assured that responses during the game would remain strictly confidential, and they were encouraged to ask questions or voice their concerns. To reduce possible lingering effects of the manipulation, participants were interviewed at length about a recent positive social experience in which they felt good.

Ethical Concerns

Ensuring children's psychological well-being during the experiment was a priority of this research. Accordingly, recommendations for experiments involving child participants that require deception and a risk of emotional discomfort were strictly followed (see Hurley & Underwood, 2002). Institutional Review Board approval was obtained, parents received a detailed consent letter describing the study, and children were preinformed about their rights to withdraw, were monitored for possible expressions of undue emotional distress, and were thoroughly debriefed. Research has shown that debriefings as described earlier are effective for preadolescent children; they typically understand how and why they were deceived, and what confidentiality entails (Hurley & Underwood, 2002). In addition, preadolescents generally are well able to regulate their emotional responses, reducing the risk of prolonged emotional discomfort (Saarni, 1999). Our own work showed that children generally are positive about participation in the Survivor Game, and report no objections against the deception or other procedural details, both immediately after the experiment and 1 week later (Reijntjes et al., 2006, 2007).

Results

Preliminary Analyses

Table 1 presents the descriptive statistics for the study variables and the results of the preliminary analyses.

Sex differences and age. Analyses of variance (ANOVAs) showed that boys reported higher state self-esteem than girls at T2 ($p < .02$) and T3 ($p < .02$), but not at T1 ($p > .07$). Sex differences in state self-esteem were not large ($ds < 0.27$). Children's age was not significantly related to any of the study variables ($ps > .08$). Because there were no interactions involving sex or age, the data for boys and girls of different ages were combined for subsequent analyses.

Equivalence of Experimental Conditions

Analyses of variance showed that children's baseline state self-esteem, narcissism levels, self-perceived social acceptance, peer-perceived social

Table 1
Descriptive Statistics for the Study Variables and Results of the Preliminary Analyses

	Disapproval (N = 108)	Control (N = 115)	Approval (N = 110)	Equivalence of conditions		Sex differences		Age	
Variable	M (SD)	M (SD)	M (SD)	F(1, 330)	p value	F(1, 331)	p value	r	p value
State self-esteem (T1)									
Total	18.43 (2.90)	18.81 (3.37)	18.61 (2.97)	0.42	> .65	3.19	> .07	−.08	> .15
Boys	18.52 (3.03)	19.00 (2.71)	19.21 (2.43)	0.89	> .41			−.10	> .24
Girls	18.36 (2.81)	18.63 (3.92)	17.96 (3.36)	0.54	> .58			−.09	> .27
Narcissism									
Total	8.59 (5.24)	8.76 (4.83)	8.70 (4.42)	0.01	> .98	3.37	> .06	.10	> .08
Boys	9.36 (5.36)	9.07 (4.66)	9.22 (4.39)	0.05	> .95			.01	> .90
Girls	8.08 (5.17)	8.38 (4.99)	8.26 (4.52)	0.06	> .94			.18	< .04
Self-perceived social acceptance									
Total	12.34 (3.07)	12.37 (3.16)	11.83 (3.41)	0.97	> .37	0.98	> .32	.01	> .91
Boys	12.22 (3.84)	12.13 (3.50)	11.69 (3.55)	0.36	> .69			−.01	> .87
Girls	12.43 (2.71)	12.59 (2.80)	11.98 (3.27)	0.64	> .52			.04	> .58
Peer-perceived social preference									
Total	0.09 (0.83)	0.02 (1.09)	−0.10 (1.05)	0.98	> .37	3.27	> .07	−.01	> .88
Boys	0.06 (0.88)	−0.13 (1.30)	−0.22 (1.13)	0.84	> .43			−.09	> .32
Girls	0.11 (0.79)	0.15 (0.85)	0.03 (0.95)	0.27	> .76			.12	> .16
Age (years)									
Total	10.73 (0.93)	10.84 (0.96)	10.76 (0.96)	0.40	> .67	2.97	> .08		
Boys	10.81 (0.90)	11.02 (1.06)	10.81 (0.89)	0.76	> .46				
Girls	10.66 (0.96)	10.69 (0.83)	10.72 (1.02)	0.05	> .95				

preference, and age did not differ in the disapproval, control, and approval groups (*p*s > .37), indicating that random assignment to conditions was successful.

Primary Analyses

Interpersonal appraisals and the modulation of state self-esteem. Data were analyzed using a mixed-model repeated measures analysis, with state self-esteem serving as the dependent variable, condition (disapproval, control, approval) serving as the between-subjects factor, and time (T1 vs. T2 state self-esteem assessment) serving as the within-subjects factor. A significant Condition × Time interaction was found, $F(2, 330) = 40.25$, $p < .001$ (see Figure 1). As predicted, disapproved children displayed a sharp decrease in state self-esteem, $F(1, 107) = 44.62$, $p < .0001$, $d = 1.29$. Control children displayed no significant change in state self-esteem, $F(1, 114) = 2.10$, $p > .14$, $d = 0.27$. Approved children displayed a sharp increase in state self-esteem, $F(1, 109) = 19.18$, $p < .0001$, $d = 0.84$. These findings show that peer approval and disapproval experiences cause immediate change in children's state self-esteem, consistent with sociometer theory.

Importantly, the effects reported earlier reflect normative, or "average" patterns of self-esteem reactivity. They provide no information on *individual differences* in self-esteem reactivity. Therefore, we computed a reliable change (RC) index (Christensen & Mendoza, 1986) to distinguish children who showed statistically significant intraindividual change in state self-esteem following the manipulated jury feedback from those who did not. The RC index provides a psychometrically sound criterion for how much change should occur between two assessments to reflect more than just random variation of an imprecise measurement instrument. Thus, it provides an ideal method to distinguish genuine self-esteem change from artifactual change (Jacobson & Truax, 1991). Using this stringent criterion, in the disapproval condition 31 children (29%) showed a significant decrease in state self-esteem. In the control condition, 4 children (4%) showed a significant increase in state self-esteem, whereas 7 children (6%) showed a significant decrease in state self-esteem. In the approval condition, 15 participants (14%) showed a significant increase in state self-esteem. Thus, in our findings substantial minorities of children who experienced peer approval or disapproval showed genuine intra-individual change in state self-esteem.

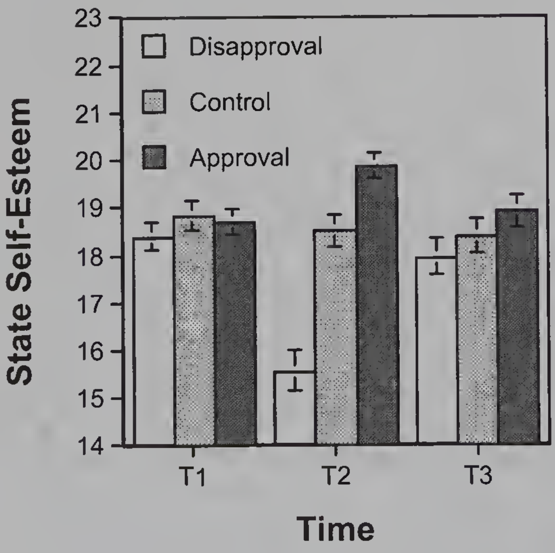


Figure 1. State self-esteem means at T1 (baseline), T2 (after jury feedback), and T3 (after viewing the individual judges' feedback) for children in the disapproval, control, and approval conditions. *Note.* Error bars represent ± 1 *SE*.

We predicted that children who showed genuine decreases in state self-esteem following disapproval would be high in narcissism but would not differ from their less reactive counterparts in self-perceived social acceptance and peer-perceived social preference. The results confirmed our expectation. An ANOVA showed that disapproved children who showed significant decreases in state self-esteem (i.e., decreases that exceeded the RC cutoff) were significantly more narcissistic ($M = 10.71$, $SD = 6.04$) than disapproved children who showed no significant decreases in state self-esteem ($M = 7.74$, $SD = 4.65$), $F(1, 106) = 7.54$, $p < .01$, $d = 0.53$. Both groups did not differ in self-perceived social acceptance ($p > .60$) or in peer-perceived social preference ($p > .47$). Similar effects were explored in the approval condition, by testing whether children who showed genuine increases in state self-esteem following approval would be high in narcissism. This was not the case. Narcissism scores were not significantly higher in approved children who showed significant increases in state self-esteem ($M = 9.40$, $SD = 5.26$) than in approved children who showed no significant increases in state self-esteem ($M = 8.59$, $SD = 4.30$), $F(1, 108) = 0.43$, $p > .51$, $d = 0.13$. Both groups did also not differ in self-perceived social acceptance ($p > .36$) and peer-perceived social preference ($p > .97$). Thus, narcissism can be considered a specific but asymmetric moderator of children's interpersonal state self-esteem reactivity, such that it is associated with increased reactivity following negative, but not positive social feedback.

Interpersonal regulation of state self-esteem. Based on sociometer theory, we predicted that peer appraisals should also influence the regulation of disapproved children's lowered state self-esteem. Specifically, positive feedback from perceived popular peers was predicted to have strong regulatory effects (Adler & Adler, 1998; Hawley et al., 2002; LaFontana & Cillessen, 2002). Data were analyzed using hierarchical regression analysis. The dependent variable was residualized state self-esteem change from T2 to T3. Residualized state self-esteem change from T1 to T2 was entered as control variable in Step 1. This statistical approach allowed us to examine regulatory change in state self-esteem (from T2 to T3) above and beyond the change that could simply be predicted from prior modulatory change in state self-esteem (from T1 to T2). The main effects for time spent viewing positive feedback from perceived popular judges and RC (i.e., a significant decrease) in state self-esteem from T1 to T2 (*no* = 0, *yes* = 1) were entered in Step 2. The two-way interaction was entered in Step 3.

The results of this hierarchical regression analysis are presented in Table 2. In Step 2 of the analysis, disapproved children showed more recovery of state self-esteem the more time they spent viewing positive feedback from perceived popular judges, $t(104) = 3.89$, $p < .001$, $b = 0.39$, $\beta = .36$. This effect, however, was qualified by its significant interaction with RC in state self-esteem from T1 to T2 in Step 3, $t(103) = 2.78$, $p < .01$, $b = 0.58$, $\beta = .29$. Despite the relatively small sample size of this group, a large and significant effect was found for disapproved children who significantly decreased in state self-esteem from T1 to T2. These children showed sharp increases in state self-esteem the more time they spent viewing positive feedback from perceived popular judges, $t(28) = 4.07$, $p < .001$, $b = 0.81$, $\beta = .57$. This relation was not significant for disapproved children who did not significantly decrease in state self-esteem, $t(74) = 1.52$, $p > .13$, $b = 0.17$, $\beta = .18$ (see Figure 2). Thus, these findings show that attending to positive feedback from perceived popular peers may help in regulating one's lowered state self-esteem, especially among children who experience marked state self-esteem losses.

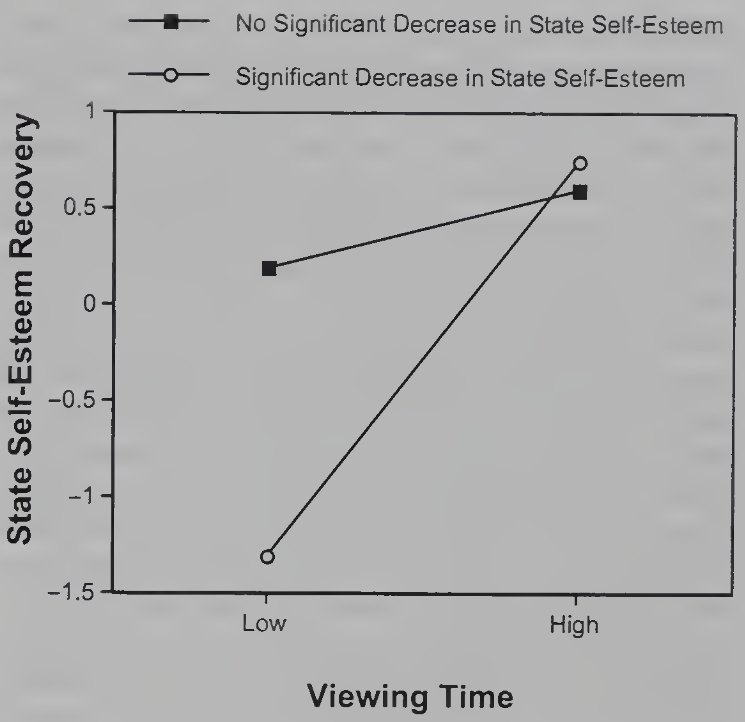


Figure 2. Relation between time spent viewing positive feedback from perceived popular judges and state self-esteem recovery for disapproved children who *did* or *did not* show a significant decrease in state self-esteem following the jury feedback.

To test the specificity of this regulatory effect of positive feedback from perceived popular judges, we conducted a series of three similar regression analyses for the other three types of feedback (i.e., positive feedback from unpopular judges, negative

Table 2

Results of Hierarchical Regression Analysis of Residualized State Self-Esteem Change From T1 to T2, the Time Children Spent Viewing Positive Feedback From Perceived Popular Judges, and Reliable Change in State Self-Esteem From T1 to T2 (No = 0, Yes = 1) *on Residualized State Self-Esteem Change From T2 to T3*

Step no.	Variable	*t*	*p* value	*b* (*SE*)	β
Step 1	Residualized state self-esteem change T1–T2	−1.84	> .06	−0.17 (0.09)	−.18
Step 2	Residualized state self-esteem change T1–T2	−2.57	< .02	−0.33 (0.13)	−.35
	Viewing time positive feedback perceived popular judges	3.89	< .001	0.39 (0.10)	.36
	Reliable change in state self-esteem	−1.24	> .21	−0.40 (0.32)	−.17
Step 3	Residualized state self-esteem change T1–T2	−2.45	< .02	−0.30 (0.12)	−.32
	Viewing time positive feedback perceived popular judges	1.82	> .07	0.21 (0.12)	.19
	Reliable change in state self-esteem	−1.29	> .20	−0.40 (0.31)	−.17
	Viewing Time × Reliable Change	2.78	< .01	0.58 (0.21)	.29

Note. $R^2 = .03$ for Step 1; $\Delta R^2 = .13$ ($p < .01$) for Step 2; $\Delta R^2 = .06$ ($p < .01$) for Step 3.

feedback from popular judges, and negative feedback from unpopular judges). No significant effects were found (*ps* > .12). Thus, the extent to which disapproved children recovered from a blow to their state self-esteem was specifically dependent on the extent to which they subsequently spent time viewing positive feedback from perceived popular judges.

If it benefits disapproved children to attend to positive feedback from perceived popular peers, one might wonder whether they also have a preference for attending to this kind of feedback over other types of feedback. Results are presented in Table 3. As expected, disapproved children spent more time viewing positive feedback from perceived popular judges than any other type of feedback (*ps* < .0001). Importantly, however, the subset of disapproved children who showed significant decreases in state self-esteem did not spend more time viewing positive feedback from perceived popular judges than did disapproved children who showed no significant decreases in state self-esteem, $F(1, 106) = 0.04$, $p > .84$, $d = 0.04$. In addition, both groups showed no different viewing times for any of the other types of feedback (*ps* > .44). Thus, although disapproved children were inclined to view positive feedback from perceived popular judges over other types of feedback, we found no evidence that their feedback viewing patterns were associated with the amount of state self-esteem they had lost.

Discussion

The present experiment found strong support for sociometer theory's core postulate that interpersonal appraisals can have an immediate, causal effect on state self-esteem. First, we found that both peer approval and disapproval experiences *modulate* children's state self-esteem. In fact, the effect size for peer approval (i.e., $d = 0.84$) exceeded Cohen's (1988) conventional value for a "large" effect (i.e., $d = 0.80$), and the effect size for peer disapproval (i.e., $d = 1.29$) was even larger. Second, we found that peer appraisals can also function to *regulate* children's state self-esteem. Specifically, the recovery of disapproved children's state self-esteem was dependent on the amount of time they spent viewing positive feedback from perceived popular peers. In contrast, the amount of time they spent viewing positive feedback from unpopular peers had no impact on their state self-esteem.

These results provide the first empirical evidence that short-term sociometer processes already operate in preadolescence, a time of great concern about others' approval (Damon & Hart, 1988; Harter, 1999, 2006; Rosenberg, 1979). At this age, children have just acquired the ability to evaluate their global worth as a person. Cognitive advances in the ability to appreciate others' appraisals in this same age period allow interpersonal experiences to have a direct impact on self-esteem (Harter, 1999, 2006). Prior longitudinal research already found that preadolescents' trait self-esteem tends to be influenced by cumulative peer experiences they have had in the past (e.g., Boivin et al., 2001; Ladd & Troop-Gordon, 2003). Our results show that preadolescents' state self-esteem is a highly dynamic psychological entity that can be immediately reactive to the social context as it currently unfolds.

The present experiment also extends the literature by examining the role of individual differences

Table 3

Time Spent by Disapproved Children on Different Types of Feedback

Viewing time in seconds (total viewing time = 300 s)	Range	All disapproved children (regardless of state self-esteem decrease; *N* = 108)	Disapproved children showing significant decrease of state self-esteem (*N* = 31)	Disapproved children showing no significant decrease of state self-esteem (*N* = 77)
		M (*SD*)	*M* (*SD*)	*M* (*SD*)
Positive feedback from perceived popular judges	11–252	94 (39)	93 (40)	95 (39)
Positive feedback from perceived unpopular judges	0–173	76 (31)	79 (35)	75 (30)
Negative feedback from perceived popular judges	0–175	66 (33)	68 (34)	65 (33)
Negative feedback from perceived unpopular judges	0–163	63 (33)	59 (38)	65 (31)

in sociometer processes. Prior research in adults primarily approached sociometer processes as universal (Denissen, Penke, Schmitt, & Van Aken, 2008; Leary & Downs, 1995; Leary et al., 1998; but see Srivastava & Beer, 2005, for a notable exception). In our study, the strength of sociometer effects varied substantially between participants. Only subsets of participants showed significant intra individual change following approval and disapproval when judged by the RC criterion (Christensen & Mendoza, 1986). It should be emphasized that this criterion is stringent. Still, this finding may raise the question whether sociometer processes are as prevalent in preadolescent children as they are in adults (Leary et al., 2003). For example, one may speculate that although preadolescents typically have acquired the cognitive abilities that allow interpersonal experiences to influence their self-esteem, these abilities are perhaps not yet as fully ingrained and habitually employed as they are later in development. It would be premature to draw such a conclusion based on our findings, but clearly, establishing the generality of sociometer processes in these relatively early stages of development will be an important direction for future research.

We predicted that the strongest sociometer effects would be found among those children who are most concerned about obtaining approval from others. Consistent with that notion, we found that children who showed marked state self-esteem losses following disapproval were high on narcissism—a trait that reflects children's preoccupation with being valued and admired by others (Morf & Rhodewalt, 2001; Thomaes, Stegge, et al., 2008). Narcissists constantly seek external self-affirmation, possibly in order to create desired self-views or to meet self-evaluative needs (Morf & Rhodewalt, 2001). Our results suggest that when they do not succeed, they immediately lose state self-esteem. Importantly, the subset of disapproved children who showed marked state self-esteem losses did not differ from others in self-perceived social acceptance and peer-perceived social preference. Thus, consistent with sociometer theory, these results suggest that interpersonal state self-esteem reactivity is most pronounced in children with a heightened need for external validation, and not just a default characteristic of children holding social vulnerabilities.

Perhaps most important was our finding that children's state self-esteem reactivity is influenced not only by the traits of the person being evaluated but also by the traits of the evaluator. In preadolescence, the perceived popularity of peers is a highly salient dimension along which children judge each other. The social power of perceived popular children reaches so far that they can dictate to a large extent what members of a social group are valued and accepted and what members are not (Adler & Adler, 1998; Hawley et al., 2002; LaFontana & Cillessen, 2002). In this light, it should come as no surprise that the manipulated popularity of judges who gave positive feedback strongly influenced the extent to which children recovered from prior disapproval. Being approved by a popular peer is significant, because it implicitly conveys the prospect that one is likely to be valued by peers in the future. Being approved by an unpopular peer is much less significant, because it tells little about how one will be valued by peers in the future.

One might be inclined to infer from these findings that disapproved children intentionally sought positive feedback from popular peers in an effort to regulate their lowered self-esteem. Although we did not test children's intentional feedback-seeking strategies (rather, we simply recorded how much time they spent viewing different types of feedback), our findings are inconsistent with such an inference. If disapproved children engaged in intentional self-esteem regulation, one would expect their feedback viewing patterns to depend on the amount of self-esteem they had lost (i.e., one would expect that disapproved children who lost the most self-esteem would spend the most time viewing positive feedback from popular peers). We found no such effect, however. Thus, although attending to positive feedback from popular peers has pronounced self-esteem regulatory effects, we have no indications that disapproved children intentionally seek such feedback to regulate their lowered self-esteem.

Limitations and Future Research

We have argued that preadolescence is a critical age period for the present experiment. However, our developmental focus limits the ability to make generalizations to children of other ages. Future research should test how interpersonal self-esteem reactivity is influenced by age. For example, is preadolescence indeed the earliest developmental period in which interpersonal self-esteem reactivity can be observed? Is interpersonal self-esteem reactivity stronger in developmental periods marked by enhanced interpersonal concern (i.e., preadolescence to middle adolescence) than in other developmental periods? Answers to such questions will help us further understand the interpersonal processes that underlie children's self-esteem development.

Laboratory experiments are the method of choice if one wants to test causal hypotheses about psychological processes. Within the confines of the laboratory, we tried to create an ecological valid peer appraisal procedure by having children interact with (ostensible) peers through the Survivor Game on the Internet. Preadolescents generally are very familiar with using social networking facilities on the Internet to communicate with peers (Valkenburg & Peter, 2007). Future research is needed to ensure that our findings generalize to other contexts of peer interaction.

It is unlikely that demand characteristics or social desirability effects have influenced the present study findings. No participants expressed suspicion about the study purposes, even when explicitly asked during the debriefing. Moreover, even if some participants had inferred the study purposes, this could not explain the interaction effects we observed (e.g., this could not explain why narcissists would be more prone than their peers to report losing self-esteem following disapproval). Still, future research focusing on more implicit measures of self-esteem (e.g., Greenwald & Farnham, 2000) may be valuable to corroborate the present study findings.

This study was conducted in the Netherlands. The determinants of self-esteem tend to be very similar across individualistic countries (Markus & Kitayama, 1991; Oyserman, Coon, & Kemmelmeier, 2002), suggesting that our findings should be generalizable to children living in North America and most Western European countries. Theoretically, sociometer effects should even be stronger in individuals from collectivistic cultures (e.g., East Asian cultures), who tend to view and evaluate themselves more in terms of their relationships with others (Oyserman et al., 2002). Future research is needed to test this hypothesis.

Our findings are at odds with the single study that examined preadolescents' actually experienced, approval-based self-esteem fluctuations (Nesdale & Lambert, 2007). One possible explanation is that Nesdale and Lambert (2007) used pretend-play peer appraisal procedures, whereas we led children to believe that they were actually appraised by others. Another possible explanation is that Nesdale and Lambert measured potential change in children's trait self-esteem, not in their state self-esteem. Continued research on the boundary conditions of the sociometer effect will enhance our understanding of the interpersonal nature of children's self-esteem.

Conclusion

This study found that children's state self-esteem can wax and wane as a function of how much they are liked by peers in the immediate present. Our findings are consistent with the view that state self-esteem functions as a sociometer, a psychological gauge of how much one is currently valued and accepted by others. As is a common characteristic of other human monitoring mechanisms (e.g., hunger), the sensitivity of the sociometer varies substantially between individuals. Just as some individuals experience changes in hunger and satiation more readily than others, some individuals experience changes in state self-esteem more readily than others. Just as some types of food alleviate hungry feelings more effectively than others, some types of social feedback seem to alleviate lowered state self-esteem more effectively than others. When children's interpersonal gauge of state self-esteem is low, praise from popular peers is likely to make it go up again.

References

Adler, P. A., & Adler, P. (1998). *Peer power: Preadolescent culture and identity*. New Brunswick, NJ: Rutgers University Press.

Ainsworth, M. D. S. (1989). Attachments beyond infancy. *American Psychologist*, *44*, 709–716.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117*, 497–529.

Baumeister, R. F., & Vohs, K. D. (2001). Narcissism as addiction to esteem. *Psychological Inquiry*, *12*, 206–209.

Boivin, M., Hymel, S., & Hodges, E. V. E. (2001). Toward a process view of peer rejection and harassment. In J. Juvonen & S. Graham (Eds.), *Peer harassment in school: The plight of the vulnerable and victimized* (pp. 265–289). New York: Guilford.

Burwell, R., & Shirk, S. (2006). The role of self-worth contingencies in adolescent depression. *Journal of Research on Adolescence, 16*, 479–490.

Bushman, B. J., & Baumeister, R. F. (1998). Threatened egotism, narcissism, self-esteem, and direct and displaced aggression: Does self-love or self-hate lead to violence? *Journal of Personality and Social Psychology, 75*, 219–229.

Christensen, L., & Mendoza, J. L. (1986). A method of assessing change in a single subject: An alteration of the RC index. *Behavior Therapy, 17*, 305–308.

Cillessen, A. H. N., & Mayeux, L. (2004). From censure to reinforcement: Developmental changes in the association between aggression and social status. *Child Development, 75*, 147–163.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Coie, J. D., Dodge, K. A., & Coppotelli, H. (1982). Dimensions and types of social status: A cross-age perspective. *Developmental Psychology, 18*, 557–570.

Cole, D. A., Jacquez, F. M., & Maschman, T. L. (2001). Social origins of depressive cognitions: A longitudinal study of self-perceived competence in children. *Cognitive Therapy and Research, 25*, 377–395.

Cooley, C. H. (1902). *Human nature and the social order*. New York: Scribner's.

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information processing mechanisms in children's social adjustment. *Psychological Bulletin, 115*, 74–101.

Damon, W., & Hart, D. (1988). *Self-understanding in childhood and adolescence*. New York: Cambridge University Press.

Denissen, J. A., Penke, L., Schmitt, D. P., & Van Aken, M. A. G. (2008). Self-esteem reactions to social interactions: Evidence for sociometer mechanisms across days, people, and nations. *Journal of Personality and Social Psychology, 95*, 181–196.

Felson, R. B., & Zielinski, M. A. (1989). Children's self-esteem and parental support. *Journal of Marriage and the Family, 51*, 727–735.

Greenwald, A. G., & Farnham, S. D. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology, 79*, 1022–1038.

Harter, S. (1985). *The Self-Perception Profile for Children: Revision of the Perceived Competence Scale for Children (manual)*. Denver, CO: University of Denver.

Harter, S. (1988). *The Self-Perception Profile for Adolescents (manual)*. Denver, CO: University of Denver.

Harter, S. (1999). *The construction of the self: A developmental perspective*. New York: Guilford.

Harter, S. (2006). The self. In W. Damon & R. M. Lerner (Series Eds.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (pp. 505–570). New York: Wiley.

Harter, S., Stocker, C., & Robinson, N. S. (1996). The perceived directionality of the link between approval and self-worth: The liabilities of a looking glass self-orientation among young adolescents. *Journal of Research on Adolescence, 6*, 285–308.

Harter, S., & Whitesell, N. R. (2003). Beyond the debate: Why some adolescents report stable self-worth over time and situation, whereas others report changes in self-worth. *Journal of Personality, 71*, 1027–1058.

Hawley, P. H., Little, T. D., & Pasupathi, M. (2002). Winning friends and influencing peers: Strategies of peer influence in late childhood. *International Journal of Behavioral Development, 26*, 466–473.

Hurley, J. C., & Underwood, M. K. (2002). Children's understanding of their research rights before and after debriefing: Informed assent, confidentiality, and stopping participation. *Child Development, 73*, 132–143.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*, 12–19.

James, W. (1890). *The principles of psychology*. New York: Dover.

Johnson, S. L., Meyer, B., Winett, C., & Small, J. (2000). Social support and self-esteem predict changes in bipolair depression but not mania. *Journal of Affective Disorders, 58*, 79–86.

Kovacs, M. (1981). Rating scales to assess depression in school-aged children. *Acta Paedo Psychiatra, 46*, 305–315.

Ladd, G., & Troop-Gordon, W. (2003). The role of chronic peer difficulties in the development of children's psychological adjustment problems. *Child Development, 74*, 1344–1367.

LaFontana, K. M., & Cillessen, A. H. N. (2002). Children's perceptions of popular and unpopular peers: A multimethod assessment. *Developmental Psychology, 38*, 635–647.

La Greca, A. M., & Stone, W. L. (1993). Social Anxiety Scale for Children–Revised: Factor structure and concurrent validity. *Journal of Clinical Child Psychology, 22*, 17–27.

Leary, M. R. (1999). Making sense of self-esteem. *Current Directions in Psychological Science, 8*, 32–35.

Leary, M. R., & Baumeister, R. F. (2000). The nature and function of self-esteem: Sociometer theory. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 32, pp. 1–62). San Diego: Academic Press.

Leary, M. R., Cottrell, C. A., & Phillips, M. (2001). Deconfounding the effects of dominance and social acceptance on self-esteem. *Journal of Personality and Social Psychology, 81*, 898–909.

Leary, M. R., & Downs, D. L. (1995). Interpersonal functions of the self-esteem motive: The self-esteem system as a sociometer. In M. H. Kernis (Ed.), *Efficacy, agency, and self-esteem* (pp. 123–144). New York: Plenum.

Leary, M. R., Gallagher, B., Fors, E., Buttermore, N., Baldwin, E., Kennedy, K., et al. (2003). The invalidity of disclaimers about the effects of social feedback on self-esteem. *Personality and Social Psychology Bulletin, 29,* 623–636.

Leary, M. R., Haupt, A. L., Strausser, K. S., & Chokel, J. T. (1998). Calibrating the sociometer: The relationship between interpersonal appraisals and state self-esteem. *Journal of Personality and Social Psychology, 74,* 1290–1299.

Leary, M. R., Tambor, E. S., Terdal, S. K., & Downs, D. L. (1995). Self-esteem as an interpersonal monitor: The sociometer hypothesis. *Journal of Personality and Social Psychology, 68,* 518–530.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 20,* 568–579.

Marsh, H. W., Craven, R. G., & Debus, R. L. (1998). Structure, stability, and development of young children's self-concepts: A multicohort-multioccasion study. *Child Development, 69,* 1030–1053.

McAninch, C. B., Manolis, M. B., Milich, R., & Harris, M. J. (1993). Impression formation in children: Influence of gender and expectancy. *Child Development, 64,* 1492–1506.

McElhaney, K. B., Antonishak, J., & Allen, J. P. (2008). "They like me, they like me not": Popularity and adolescents' perceptions of acceptance predicting social functioning over time. *Child Development, 79,* 720–731.

Mead, G. H. (1934). *Mind, self, and society.* Chicago: University of Chicago Press.

Morf, C. C., & Rhodewalt, F. (2001). Unraveling the paradoxes of narcissism: A dynamic self-regulatory processing model. *Psychological Inquiry, 12,* 177–196.

Mrug, S., & Hoza, B. (2007). Impression formation and modifiability: Testing a theoretical model. *Merrill-Palmer Quarterly, 53,* 631–659.

Nesdale, D., & Lambert, A. (2007). Effects of experimentally manipulated peer group rejection on children's negative affect, self-esteem, and maladaptive social behavior. *International Journal of Behavioral Development, 31,* 115–122.

Oosterwegel, A., Field, N., Hart, D., & Anderson, K. (2001). The relation of self-esteem variability to emotion variability, mood, personality traits, and depressive tendencies. *Journal of Personality, 69,* 689–708.

Oyserman, D., Coon, H. M., & Kemmelmeier, M. (2002). Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin, 128,* 3–72.

Patterson, C. J., Kupersmidt, J. B., & Griesler, P. C. (1990). Children's perceptions of self and of relationships with others as a function of sociometric status. *Child Development, 61,* 1335–1349.

Raskin, R., & Terry, H. (1988). A principal-components analysis of the Narcissistic Personality Inventory and further evidence of its construct validation. *Journal of Personality and Social Psychology, 54,* 890–902.

Reijntjes, A., Dekovic, M., Vermande, M., & Telch, M. J. (2007). Children's feedback preferences in response to an experimentally manipulated peer evaluation outcome: The role of depressive symptoms. *Journal of Abnormal Child Psychology, 35,* 497–507.

Reijntjes, A., Stegge, H., Meerum Terwogt, M., Kamphuis, J. H., & Telch, M. J. (2006). Emotion regulation and its effects on mood improvement in response to an in vivo peer rejection challenge. *Emotion, 6,* 543–552.

Rhodewalt, F., & Morf, C. C. (1998). On self-aggrandizement and anger: A temporal analysis of narcissism and affective reactions to success and failure. *Journal of Personality and Social Psychology, 74,* 672–685.

Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin, 128,* 934–960.

Rosenberg, M. (1965). *Society and the adolescent self-image.* Princeton, NJ: Princeton University Press.

Rosenberg, M. (1979). *Conceiving the self.* New York: Basic Books.

Rubin, K. H., Bukowski, W. M., & Parker, J. G. (2006). Peer interactions, relationships, and groups. In W. Damon & R. M. Lerner (Series Eds.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (pp. 571–645). New York: Wiley.

Rudolph, K. D., Caldwell, M. S., & Conley, C. S. (2005). Need for approval and children's well-being. *Child Development, 76,* 309–323.

Saarni, C. (1999). *The development of emotional competence.* New York: Guilford.

Salmivalli, C., & Isaacs, J. (2005). Prospective relations among victimization, rejection, friendlessness, and children's self-and peer-perceptions. *Child Development, 76,* 1161–1171.

Srivastava, S., & Beer, J. S. (2005). How self-evaluations relate to being liked by others: Integrating sociometer and attachment perspectives. *Journal of Personality and Social Psychology, 89,* 966–977.

Thomaes, S., Bushman, B. J., Orobio de Castro, B., Cohen, G. L., & Denissen, J. J. A. (2009). *Reducing narcissistic aggression by buttressing self-esteem: An experimental field study.* Manuscript submitted for publication.

Thomaes, S., Bushman, B. J., Stegge, H., & Olthof, T. (2008). Trumping shame by blasts of noise: Narcissism, self-esteem, shame, and aggression in young adolescents. *Child Development, 79,* 1792–1801.

Thomaes, S., Stegge, H., Bushman, B. J., Olthof, T., & Denissen, J. (2008). Development and validation of the Childhood Narcissism Scale. *Journal of Personality Assessment, 90,* 382–391.

Valkenburg, P. M., & Peter, J. (2007). Preadolescents' and adolescents' online communication and their closeness to friends. *Developmental Psychology, 43,* 267–277.

Veerman, J. W., Straathof, M. A. E., Treffers, P. D. A., Van den Bergh, B. R. H., & ten Brink, L. T. (1997). *Han-*

dleiding Competentiebelevingsschaal voor Kinderen CBSK [Manual for the Dutch version of the Self-perception profile for children]. Lisse, Netherlands: Swets & Zeitlinger.

Appendix: State Self-Esteem Scale

1. I am satisfied with myself right now.
2. I feel down on myself right now.*
3. I am proud of myself right now.
4. I am dissatisfied with myself right now.*
5. I feel good about who I am right now.
6. I am disappointed in myself right now.*

Note. Responses are scored using a 5-point scale ranging from 0 (*not at all*) to 4 (*very much*). Items marked with an asterisk were reverse scored.

Child Development, May/June 2010, Volume 81, Number 3, Pages 826–836

Individual and Peer Group Normative Beliefs About Relational Aggression

Nicole E. Werner and Laura G. Hill
Washington State University

Studies show that children who use relational aggression process social information in unique ways; however, findings have been inconsistent and limited by methodological weaknesses. This short-term longitudinal study examined developmental changes in 245 (49% female; ages 8–13) 3rd through 8th graders' normative beliefs about relational aggression and tested the hypothesis that individual and classroom-level norms predict relational aggression 1 year later. Results showed that the transition to middle school was marked by increased approval of relational aggression, and individual norms predicted future relational aggression. Importantly, a contextual model showed that students in peer groups highly supportive of relational aggression became increasingly aggressive. Findings extend social information processing theories of relational aggression to focus on the role of peer group cognitions.

Empirical support for social information processing (SIP) models of children's social behavior suggests that patterns of both distal (i.e., latent knowledge structures) and proximal (i.e., online) processing of social information predict individual differences in overt aggression (see Crick & Dodge, 1994, for a review). Although some evidence suggests that SIP mechanisms may be important for understanding the development and maintenance of *relational aggression*—a form of aggression characterized by actual or threatened damage to relationships (e.g., Crick & Grotpeter, 1995), results of prior work in this area have been mixed (e.g., Crain, Finch, & Foster, 2005; MacBrayer, Milich, & Hundley, 2003; Werner & Nixon, 2005). The current study was designed to further examine SIP factors and relational aggression by focusing on children's normative beliefs. This study extends previous work by exploring developmental differences in children's beliefs about relational aggression and by examining the influence of peer-group-level normative beliefs in the prediction of relational aggression.

SIP and Relational Aggression

SIP models of children's social adjustment suggest that aggressive behavior results from deficient or biased online processing of social information (Crick & Dodge, 1994), and a large body of literature links overtly aggressive behavior with maladaptive SIP patterns. In contrast, only a few studies of SIP and relational aggression have been published to date. Crick and colleagues (Crick, 1995; Crick, Grotpeter, & Bigbee, 2002) found that children who use relational aggression at high levels attribute hostile intent to peers in ambiguous relational conflicts (i.e., involving threats to their acceptance by friends or the peer group) whereas physically aggressive children's hostile attribution biases are limited to instrumental provocation situations. Although promising, Crick and colleagues' findings have not been replicated (cf. Crain et al., 2005; Nelson, Mitchell, & Yang, 2008). Moreover, several studies of other SIP steps (e.g., goal clarification, response evaluation) have failed to find that relationally aggressive children display unique SIP patterns (Crick & Werner, 1998; Delveaux & Daniels, 2000).

Normative Beliefs About Aggression

One component of the SIP model that has received increasing attention is the "database." SIP theory posits that knowledge structures function as a cross-situational, distal storehouse of information that controls children's behavior by imposing limits on the processing of specific information in the social environment (i.e., online

---

Correspondence concerning this article should be addressed to Nicole E. Werner, Washington State University, 501 Johnson Tower, Department of Human Development, Pullman, WA 99164-4852. Electronic mail may be sent to nwerner@wsu.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8103-0012

processing). Huesmann (1988) and Crick and Dodge (1994) argued that children's memories of past experiences are stored in long-term memory and, over time, integrated into generalized knowledge structures such as scripts, schemas, or working models (Dodge, 1993). Knowledge structures simplify the task of processing cues in the immediate social environment, resulting in more efficient, but not necessarily more accurate or adaptive, processing.

One line of work on latent knowledge structures has explored children's generalized beliefs about the legitimacy or normative nature of aggression. Converging evidence suggests that children who view aggression as an acceptable response are more likely to engage in aggressive behavior concurrently and in the future (Henry et al., 2000; Huesmann & Guerra, 1997). Several studies have explored children's normative beliefs about relational aggression. Musher-Eizenman et al. (2004) surveyed 778 children in Grades 4–6 and found that those children who were more approving of indirect aggression (a construct similar to relational aggression) reported higher levels of indirect aggression. Similar findings were reported by Werner and Nixon (2005) in a study of sixth through eighth graders, and by Bailey and Ostrov (2008) in a study of college students. Only one study has examined children's normative beliefs and relational aggression using an independent informant of aggression. Murray-Close, Crick, and Galotti (2006) found that fourth and fifth graders' judgments of the appropriateness of relational aggression were marginally associated with peer reports of relational aggression.

In sum, prior research has yielded some evidence to support the application of the SIP model to children's relational aggression. However, the failure to replicate some findings has led some researchers to call into question the utility of the SIP model for understanding relational aggression (Crain et al., 2005; Nelson et al., 2008). Methodological limitations, including the reliance of many studies on self-reports of aggression and cross-sectional designs, might explain, in part, the lack of compelling findings. It is also possible that the sole focus on *individual* social cognitive processes in prior research has limited our understanding of the significance of SIP for relational aggression. In this study, we hypothesized that social cognitions at the level of the *peer group* will provide important information about the development and/or maintenance of children's relational aggression.

The Role of the Peer Group for Relational Aggression

A growing body of literature points to the significant role that peer relationships play for relational aggression during childhood and adolescence. Several studies have found that close friends tend to resemble each other in levels of relational aggression (Duffy & Nesdale, 2008; Ellis & Zarbatany, 2007; Werner & Crick, 2004), and having relationally aggressive friends predicts increases in children's relational aggression over time (Werner & Crick, 2004). Socialization processes at the dyadic and larger peer group level might explain these findings. Werner and Crick (2004) hypothesized that relational aggression is selectively reinforced within the close friendships of children who exhibit relationally aggressive behaviors through a process similar to deviancy training that characterizes the friendships of antisocial males (Dishion, Patterson, & Griesler, 1994). Consistent with this notion, Cillessen and Mayeux (2004) interpreted findings of strong associations between relational aggression and perceived popularity as indicating that relational aggression is increasingly reinforced in the peer group with age, via the benefits incurred from high status.

The previously described findings raise the possibility that peer group norms, such as the level of acceptance of aggression, may play an important role in the development and maintenance of relationally aggressive behaviors. The literature on bullying has long recognized the impact of group processes, although operational definitions of group norms have varied widely across studies. Salmivalli and Voeten (2004) measured students' expectations of the social consequences of bullying in their classroom and found that group norms explained classroom variance in bullying behavior above and beyond individuals' beliefs about the appropriateness of bullying. Duffy and Nesdale (2008) defined group norms as students' ratings of "how happy" their peer group would be if one of its members bullied. Students in peer groups more supportive of bullying reported engaging in higher levels of bullying themselves.

In this study, we hypothesized that both individual beliefs about the acceptability of relational aggression, and those of the larger peer group (i.e., the classroom), influence students' engagement in relationally aggressive behavior. In the most direct test of this hypothesis, Henry et al. (2000) examined the roles of individual and classroom-level normative beliefs in predicting individual overt aggression and norms 2 years later in a sample of urban second through sixth graders. Classroom norms were positively associated with sixth graders' individual levels of aggression after controlling for earlier levels of aggression. In addition to this direct effect, an indirect effect of group norms was found for the whole sample, such that classroom norms predicted individual norms, which in turn were associated with individual aggression.

Another way to conceptualize group norms for aggression is to consider levels of aggression in the peer group, sometimes referred to as descriptive norms. Espelage, Holt, and Henkel (2003) identified peer groups within middle schools and gathered self-reports of bullying and fighting behavior. Peer group levels of bullying predicted individual levels of bullying assessed 6 months later after controlling for the stability of individual bullying. In a recent study, Kuppens, Grietens, Onghena, Michiels, and Subramanian (2008) found that third- through fifth-grade children in classrooms characterized by higher relational aggression became increasingly relationally aggressive across a 1-year period. Together, these findings support the hypothesis that normative (i.e., frequent) behavior in the peer group influences individual children's behavior via modeling or peer socialization processes (Bandura, 1986).

The current investigation was designed to increase our understanding of SIP factors in relational aggression by exploring the independent and interactive effects of individual- and classroom-level normative beliefs for children's relationally aggressive behaviors. Methodologically, this study extends prior research about relational aggression in two ways: (a) we used a short-term longitudinal design and studied a wide range of grades, and (b) relational aggression was assessed using an independent informant (teachers).

Our first goal was to examine developmental differences in normative beliefs about relational aggression, a previously unexplored area. Zelli, Dodge, Lochman, and Laird (1999) found that mean levels of normative beliefs changed little from third through fifth grade and that individual differences in beliefs were moderately stable. van Schoiack-Edstrom, Frey, and Beland (2002) collected pre- and posttest data on middle school students' beliefs about social exclusion in schools receiving a prevention program (Second Step©) and comparison schools. Students in the comparison schools became increasingly approving of social exclusion over the course of one semester, relative to students in the intervention schools who decreased their approval of social exclusion. Given the lack of previous work on the development of normative beliefs about relational aggression, our predictions should be considered exploratory. We hypothesized that children become increasingly approving of relational aggression across childhood and adolescence, whereas approval of overt aggression either remains stable or decreases (cf. Zelli et al., 1999). Because our longitudinal design included children who transitioned from elementary to middle school during the study, we also examined whether this developmental transition is marked by changes in approval of both forms of aggression, as might be predicted by Pellegrini and Bartini (2001).

Our second goal pertained to the role of individual and peer group normative beliefs in predicting changes in relational aggression. Consistent with previous studies of overt aggression (e.g., Henry et al., 2000; Huesmann & Guerra, 1997), we expected that both individual and group norms about relational aggression would explain significant variance in relationally aggressive behavior. We further hypothesized that individual and group norms would interact, such that the highest levels of relational aggression were expected among children who approved of relational aggression *and* who were a part of peer groups in which relational aggression was viewed positively. We also tested a reciprocal model whereby students' level of relational aggression predicted change over time in normative beliefs. Drawing on the work of Huesmann and Guerra (1997), we expected to find that for children in this age group, aggression would not explain significant variance in individual norms.

Method

Participants

Participants in the current study were drawn from a 3-year longitudinal study of school climate. The 2nd-year sample (referred to here as Time 1 or T1) included third through eighth graders and their teachers from all three elementary schools and one middle school in a midsized town in the Western United States. A total of 726 students completed the survey (53% female and 73% of the population of third through eighth graders). We did not collect information about student ethnicity or race. However, the population from which this sample was drawn consisted of 77% White, 5% Black, 12% Asian, and 4% Hispanic students. Students were fairly evenly distributed across Grades 3 through 8, with slightly more than half of participants in elementary school (third–fifth grades). Teachers completed ratings for 523 of participating students.

Elementary school teachers were more likely to complete ratings than were middle school teachers ($\chi^2 = 82.1$, $p < .001$). Comparison of students with and without teacher ratings on normative beliefs and aggression variables revealed no significant differences.

The following year, students in Grades 3 through 8 were again recruited for participation. We did not have permission to collect data in the single local high school; thus, we were unable to recruit T1 eighth graders who transitioned to the ninth grade the next year. At Time 2 (T2), 467 students repeated the survey (64% of the T1 sample), and teachers completed ratings for 305 of these students. Again, comparison of students with and without teacher ratings revealed no significant differences on variables of interest in this study.

Procedure

Consent forms were sent home with all third- through eighth-grade children during the 1st year of the study. Parents were required to give active consent for their child to participate in this research. The following year, a passive consent process was used. Parents of participating children from the prior year received a letter from the school reminding them of the study and indicating that they should contact the school if they did not want their child to participate. Consent forms were sent home with all children who were new to the school and who had not received a permission form in the previous year. Before administering surveys to children, children were read an assent statement and asked to sign a form indicating assent.

In both years, surveys were administered by trained research assistants and the study authors during a 1-week period in the middle of fall semester. In each classroom, one person read survey items aloud while the other canvassed the room to assist students who needed help. Data collection took approximately half an hour in each classroom. Absent students were surveyed in groups during the week after initial administration.

Teachers completed survey instruments assessing child adjustment during the same semester. Elementary school children's primary teacher completed assessments. For students in middle school, we selected fourth-period teachers to complete the ratings because students spent the equivalent of two class periods in fourth period. We therefore reasoned that these teachers would have the most knowledge of students' social behavior. All study procedures were approved by the Institutional Review Board of the university where the authors were employed.

Measures

Teacher ratings of aggression. Teacher ratings of child aggression were obtained using two measures validated in prior research. Relational aggression was assessed with five items from the Children's Social Behavior Scale–Teacher Form (Crick, 1996). Sample items included, "when this child is mad at a peer, s/he gets even by excluding the peer from the group" and "this child spreads rumors or gossips about some peers." Each item was rated on a 5-point scale (*never true* to *almost always true*). Crick (1996) reported that the scale emerged as a separate factor from overt aggression and prosocial behavior in a principal components factor analysis (PCFA), and the alpha was .94 in a sample of third through sixth graders. The scale also correlated with peer reports of relational aggression, .57 for boys and .63 for girls. Overt aggression was assessed with three items from the Teacher Observation of Classroom Behavior–Revised (TOCA–R; Corrigan, 2003; Werthamer-Larsson, Kellam, & Wheeler, 1991). Sample items included "fights" and "harms others." Items were rated on 6-point Likert-type scales ranging from 1 (*never*) to 6 (*almost always*).

We conducted a confirmatory factor analysis using MPlus 4.21 (Muthén & Muthén, 1998–2007) on the eight aggression items. A one-factor model fit the data poorly, $\chi^2(20, N = 533) = 319.35$, $p < .001$; root mean square error of approximation (RMSEA) = .168, standardized root mean square residual (SRMR) = .094, comparative fit index (CFI) = .905. In contrast, a two-factor model fit the data significantly better, $\chi^{2\mathrm{diff}}(1) = -218.42$, $p < .001$. The final model provided an adequate fit to the data, $\chi^2(19) = 100.93$, $p < .001$; RMSEA = .09, SRMR = .049, CFI = .97. We computed scores for relational aggression ($\alpha = .94$) and overt aggression ($\alpha = .80$).

Normative Beliefs About Aggression

Normative beliefs about relational and overt aggression were assessed at T1 and T2 with items used in prior research (Huesmann & Guerra, 1997; Werner & Nixon, 2005). Participants rated the acceptability of aggressive behaviors on 5-point Likert scales ranging from 1 (*no!*) to 5 (*yes!*). Higher scores reflected greater approval of aggression. Normative beliefs about relational aggression were assessed with three items (e.g., "In general, it is OK to tell your friends not to be friends with someone you don't like''). Beliefs about overt aggression were assessed with four items (e.g., ''In general, it is OK to hit other people'').

We conducted a confirmatory factor analysis using MPlus 4.21 (Muthén & Muthén, 1998–2007) on the seven normative belief items. A one-factor model yielded good fit, $\chi^2(14) = 72.0$, $p < .001$; RMSEA = .08, SRMR = .04, CFI = .97. However, a two-factor model fit the data significantly better, $\chi^{2\text{diff}}(1) = -46.83$, $p < .001$. The final model provided an excellent fit to the data, $\chi^2(13) = 25.17$, $p = .02$; RMSEA = .04, SRMR = .02, CFI = .99. In the current study, the beliefs about relational aggression and beliefs about overt aggression scales demonstrated acceptable reliability (αs = .64 and .87, respectively).

Creation of peer group variables. To evaluate the contextual effects of peer group normative beliefs on changes over time in aggression, we calculated a mean score on the normative belief scale for each classroom and subtracted it from each individual's score (Hofmann & Gavin, 1998; Raudenbush, 1989). Thus, each student had two scores for normative beliefs about relational and overt aggression: the contextual score, a within-classroom peer-group average; and the ''pure'' individual score with peer group average subtracted.

Attrition and Missing Data

T1 to T2 attrition was 36% in this study (although this rate is somewhat misleading given that T1 eighth graders were not invited to participate in T2). Rates of 40%–50% are not uncommon in school-based longitudinal studies (Foshee et al., 2005; Horton & Lipsitz, 2001). Attrition analyses indicated that the longitudinal sample did not differ from the group of students with only Time 1 data on variables relevant for this study. Some analyses reported in this study were limited to students with complete self-report and teacher-report data from both T1 and T2 assessments ($N = 245$; 49% female). A one-way analysis of variance (ANOVA) comparing this subsample with the T1 only and T2 only samples revealed no significant differences on any of the study variables. In order to account for possible bias introduced by attrition between Times 1 and 2, we conducted analyses with multiply imputed data (Schafer, 1999) using the MI and MI-Analyze procedures available in SAS version 9 (Yuan, 2000). Significance of findings was the same with and without multiple imputation; therefore, we present all results using the original data.

Results

Developmental Differences in Individual Normative Beliefs

Our first goal was to examine developmental differences in students' normative beliefs about relational and overt aggression. We computed a repeated measures ANOVA with time and aggression type (relational norms, overt norms) as within-subjects factors, and sex and school change (elementary, middle, transition to middle) as the between-subjects factors (Table 1).

The analysis yielded several significant main effects: sex, $F(1, 459) = 6.85$, $p < .01$; school change, $F(2, 459) = 45.08$, $p < .001$; aggression type, $F(1, 459) = 5.21$, $p < .001$; and time, $F(1, 459) = 16.77$, $p < .001$. Examination of means showed that males were more approving of aggression than were females (Ms = 1.76 vs. 1.64), and students in middle school ($M = 1.92$) and those who transitioned to middle school during the study ($M = 1.72$) were more approving of aggression compared to elementary school students ($M = 1.45$). Students were more approving of relational aggression than overt aggression (Ms = 1.99 vs. 1.41), and students became increasingly approving of aggression over the 1-year study period (Ms = 1.64 vs. 1.76). The main effects of time, aggression type, and school change were qualified, however, by significant two-way and three-way interactions. For the sake of parsimony, we report only three-way interaction results and follow-up analyses.

There was a significant multivariate interaction of time, aggression type, and school change, $F(2, 459) = 7.27$, $p < .001$. To explore this effect further, we recomputed the analysis separately for students in elementary school, middle school, and those who transitioned to middle school during the study. The interaction of time and aggression type was nonsignificant for elementary students, $F(1, 208) = 1.87$, $p < .07$, but significant for middle school students, $F(1, 148) = 34.66$, $p < .001$, and transition students, $F(1, 103) = 28.53$, $p < .001$.

To probe these effects further, we recomputed the analysis of middle school and transition students' beliefs separately for relational aggression norms and overt aggression norms, thus focusing on the time effect singly. In the analysis of overt aggression norms, the effect of time was nonsignificant for both middle school students, $F(1, 148) = .04$, $p > .05$, and transition students, $F(1, 103) = 3.51$, $p < .10$. Thus, students' normative beliefs about overt aggression remained relatively stable across the 1-year period. The effect of time on relational

Table 1

Teacher Reports of Aggressive Behavior, Youth Demographics, and Self-Reports of Normative Beliefs: Correlations and Descriptive Statistics

Variable	1	2	3	4	5	6	7	8	9	10
1. Youth gender[a]	—									
2. Grade (3rd–8th)	.08*	—								
3. Overt aggression (T1)	−.17**	.08	—							
4. Overt aggression (T2)	−.12*	−.05	.42**	—						
5. Relational aggression (T1)	.07	−.03	.70**	.31**	—					
6. Relational aggression (T2)	.05	−.10	.29**	.69**	.31**	—				
7. OA norms (T1)	−.11**	.32**	.16**	.09	.12**	.05	—			
8. OA norms (T2)	−.17**	.39**	.17**	.14**	.08	.05	.44**	—		
9. RA norms (T1)	−.10**	.20**	.13**	.16**	.07	.11*	.61**	.34**	—	
10. RA norms (T2)	−.06	.41**	.06	.03	.01	.00	.26**	.55**	.43**	—
N	722	727	522	495	512	498	726	693	726	692
M	0.53	1.83	1.61	1.56	1.99	1.99	1.48	1.35	1.91	2.01
SD	0.50	0.91	0.73	0.74	0.94	0.98	0.68	0.56	0.79	0.81
Range	0–1	1–3	1–6	1–6	1–5	1–5	1–5	1–5	1–5	1–5
α	n/a	n/a	.80	.88	.94	.97	.87	.85	.64	.62

Note. RA = relational aggression; OA = overt aggression; norms = normative beliefs about aggression.

[a] Youth gender: 0 = *male*, 1 = *female*.

$^*p < .05$. $^{**}p < .005$.

aggression norms was significant for both middle school students, $F(1, 148) = 32.53$, $p < .001$, and transition students, $F(1, 143) = 30.46$, $p < .001$. Both groups of students reported increases in approval of relational aggression over time (middle school: *Ms* = 2.05 vs. 2.39; transition: *Ms* = 1.83 vs. 2.29).

Prediction of Relational Aggression From Individual Normative Beliefs (Model 1)

Hierarchical linear regression was used to test the hypothesis that students' normative beliefs about relational aggression would predict future (T2) relational aggression. The results of this analysis are presented in Table 2 (Model 1). In the first step of the model, sex and school type were entered. At the second step, we entered T1 scores for relational and overt aggression. This allowed us to: (a) predict *changes* in relational aggression by controlling for initial levels and (b) predict unique variance in relational aggression by controlling for overlap with overt aggression. At the third step we entered T1 normative beliefs as well as concurrent normative belief variables (T2; relational and overt norms). Thus, we explored whether normative beliefs about relational aggression predicted unique variance in the corresponding aggressive behavior 1 year later, after controlling both for the stability of normative beliefs and aggression and the overlap between relational and overt aggression.

The final regression model was significant, $F(8, 240) = 6.36$, $p < .001$. Importantly, variables entered at the third step showed that the addition of initial normative beliefs scores explained a significant unique proportion of variance in T2 relational aggression, $F\Delta(5, 232) = 4.85$, $p < .01$. Examination of beta weights in the final model showed that T1 relational aggression ($\beta = .26$, $p < .001$) and T1 normative beliefs about relational aggression ($\beta = .20$, $p < .01$) were the only significant predictors of T2 relational aggression scores.

To test whether these results were robust across school types, we recomputed the above analysis and included the interactions of the two school type dummy variables with normative beliefs about relational aggression at T1 at the final step. Again, the overall regression model was significant, $F(11, 240) = 5.21$, $p < .001$. The final step that included the interactions, however, was nonsignificant, $F\Delta(2, 229) = 2.96$, $p < .06$. We therefore concluded that these findings were robust across school types.

Finally, we tested a reciprocal model to rule out bidirectional effects, with relational aggression at T1 predicting normative beliefs at T2 and using the same control variables. In this model, normative beliefs at T1 predicted normative beliefs at T2 ($\beta = .30$, $p < .001$), but T1 relational aggression did not predict T2 normative beliefs ($\beta = .02$, $p < .73$).

Effects of Peer Group Normative Beliefs on Relational Aggression (Model 2)

To test the hypothesis that both individual and classroom-level norms would explain significant

Table 2
Normative Beliefs About Relational Aggression Predicting Change Over Time in Relationally Aggressive Behaviors (N = 241)

	Relational aggression Time 2					
	Model 1			Model 2		
Variable	*B*	*SE B*	β	*B*	*SE B*	β
Sex	0.16	0.12	0.08	0.22	0.12	0.11
Elementary school	−0.27	0.16	−0.14	−0.32	0.16	−0.16*
Middle school	0.25	0.19	0.10	0.00	0.20	0.00
Overt aggression Time 1	0.14	0.14	0.10	0.20	0.13	0.14
Relational aggression Time 1	0.28	0.10	0.26**	0.21	0.10	0.20*
Overt aggression norms Time 2	−0.18	0.15	−0.10	−0.19	0.14	−0.11
Relational aggression norms Time 2	−0.07	0.11	−0.06	−0.06	0.10	−0.05
Relational aggression norms Time 1	0.25	0.08	0.20**	—	—	—
Individual relational aggression norms Time 1[a]	—	—	—	−0.50	0.56	−0.36
Classroom relational aggression norms Time 1[a]	—	—	—	1.13	0.24	0.31**
Individual × Classroom Norms Time 1[a]	—	—	—	0.34	0.30	0.46
Adj. R^2		0.15			0.20	
F		6.36**			7.10**	

[a] These variables represent the decomposition of the unitary score on relational aggression normative beliefs into the two scores (individual and classroom normative beliefs) and their interaction, used in Model 2. Interactions of normative beliefs with grade level were not significant and were dropped from the model.
$^{*}p < .05$. $^{**}p < .01$.

variance in future levels of relational aggression, we recomputed the above model with three changes. First, we replaced the individual normative belief variables with the corresponding "pure" variable representing the individual norm effect minus the classroom effect. Second, we included classroom-level scores for relational aggression norms in the analysis. Finally, we computed the interaction of pure individual norms and classroom-level norms and included this variable in the analysis (refer to Table 2, Model 2 for results).

As predicted, results showed that peer group norms for relational aggression significantly predicted increases in relationally aggressive behavior over time ($\beta = .31$, $p < .001$). In other words, students in classrooms high in approval of relational aggression became increasingly aggressive across the 1-year period. Interestingly, and in contrast to the previous analysis, the relation of individual norms to later aggression was nonsignificant ($\beta = -.36$, $p < .37$) when peer-level norms were included in the model.

Consistent with our expectations, increases in levels of relational aggression were lower for elementary school students ($\beta = -.16$, $p < .05$) than for the middle school and transition comparison groups. Model fit was significantly higher than that of the model that assessed relational aggression as a unitary construct $F\Delta(2, 230) = 8.35$, $p < .01$.

Discussion

The results of this study provide important information about children's normative beliefs about relational aggression and the significance of these beliefs at the peer group level for the development of relational aggression during middle childhood and early adolescence.

Using a conservative analytic approach, we found that students who reported greater approval of relational aggression were more relationally aggressive 1 year later, even after controlling for the across-time stability of relational aggression and normative beliefs, and for the overlap between relational and overt forms of aggression. We also tested a reciprocal model and found that, consistent with Huesmann and Guerra (1997), by the late elementary school years individual differences in aggressive behavior did not significantly influence children's normative beliefs about aggression. Together, these results provide support for the SIP framework whereby normative beliefs operate as a script or schema that increases the efficiency with which the child processes social information in the immediate environment. For example, scripts should increase the likelihood that the child will encode threatening cues in peer conflict situations (e.g., a peer is standing too close to my romantic partner), make maladaptive attributions about others' intentions (e.g., a peer is threatening my social status), access aggressive responses from memory, and enact those responses. Although this study did not explore online information processing as a mediator of the link between normative beliefs and relational aggression, an important direction for future research, our findings strengthen results from previous investigations by using independent informants of relational aggression and a longitudinal design. These findings illustrate the need for continued investigation into the social-cognitive bases of relational aggression, particularly in the ''database'' of the SIP model.

Our results also demonstrated remarkable specificity in norms–behavior associations by form of aggression. Despite the moderate to high intercorrelations of aggression forms and beliefs about aggression forms within assessments, relational aggression and relational aggression norms were the only unique predictors of future relationally aggressive behavior. These results provide further evidence of the distinctiveness of relational and overt forms of aggression.

The most compelling results of the current study, however, focus on the role of the peer group for the development of relational aggression. Using a contextual analytic approach, we found that when support for relational aggression in the classroom was high, students in those classrooms were more relationally aggressive the following year compared to students in peer groups less supportive of relational aggression. Contrary to expectations, when normative beliefs were partitioned into peer-group-level and individual-level components, individual normative beliefs did not explain significant variance in, nor did they interact with group norms in the prediction of, future relational aggression.

One explanation for the significance of peer-group-level norms for individuals' aggressive trajectories in this study might lie in the link between relational aggression and peer status. In early adolescence, relationally aggressive youths are perceived by their peers as ''popular'' (Rose, Swenson, & Waller, 2004), an association that strengthens across the adolescent period (Cillessen & Mayeux, 2004). We expect that perceptions of popularity are interconnected with students' approval of the behavior exhibited by popular youth, and these linkages are likely bidirectional in nature. Observations that youth who use relational aggression confer privileged status in the peer group might contribute to more positive evaluations of the behaviors exhibited by those youth, a hypothesis that is supported by our finding that early adolescence was marked by a significant increase in approval of relational aggression and higher levels of relational aggression. At the same time, as relationally aggressive behaviors come to be viewed more positively by adolescents, perceptions of youth who engage in these behaviors should improve. Regardless of the direction of effects, the benefits of high social status incurred by youth who use relational aggression are likely to act as reinforcement of relational aggression, thus explaining the impact of group norms on individual behavior.

Another explanation for the significance of peer group normative beliefs relates to deviancy training and peer contagion. Dishion and colleges have convincingly shown that aggregating high-risk peers increases the risk for escalation of deviant behavior (e.g., Dishion, Eddy, Haas, Li, & Spraklen, 1997; Dishion, McCord, & Poulin, 1999). If group-level norms mirror levels of actual aggression in the peer group, then our findings could reflect peer contagion of relational aggression. Support for this hypothesis lies in a recent study that found variation among classrooms in levels of relational aggression explained significant variance in elementary school-aged children's relational aggression assessed 1 year later (Kuppens et al., 2008). An important direction for future research will be to tease out the influence of peer group aggression and peer group norms in the prediction of children's relational aggression, and to examine the mediating role of children's SIP patterns in this process.

Two additional findings warrant mention. First, although we did find gender differences in approval of aggression (males were more approving) and in overt aggression (males were rated higher), gender was unrelated to relational aggression. These findings are consistent with the growing literature showing that gender differences in relational aggression are generally small (Card, Stucky, Sawalani, & Little, 2008). Concerns about lack of power precluded us from testing gender-specific models; however, on the basis of past research we did not expect gender differences in the relations of normative beliefs and relational aggression (Werner & Nixon, 2005) or of classroom norms and aggression (e.g., Salmivalli & Voeten, 2004). Given previous findings of the importance of friendships for girls' relational aggression (Werner & Crick, 2004), however, it will be important for future studies to explore group norms and socialization processes within more intimate peer groups. Laboratory studies of micro-level interaction dynamics hold particular promise in this area (e.g., Underwood & Buhrmester, 2007).

Second, the developmental and grade-related effects found in this study point to the need to consider how the middle school context contributes to students' beliefs about relational aggression. As noted by other researchers, middle schools are typically characterized by a larger student body, shifting peer groups, more impersonal teacher–student relationships, and less adult monitoring (e.g., Eccles et al., 1993). These contextual factors might contribute to increases in relational aggression directly, by reducing adult sanctions against the behavior. Indirect influences, however, are also possible. The majority of school-based bullying prevention efforts emphasize overt forms of aggression. Under these circumstances, students are likely to observe that adults pay little attention to relationally aggressive behavior (relative to overt aggression) and internalize these reactions as supportive of relational aggression. These processes might explain, in part, the increases in acceptability of relational aggression and growth in relational aggression among older students in this study.

Not only do these results contribute to the theoretical and empirical literature on relational aggression, they provide important directions for prevention and intervention work. Previous work has demonstrated the malleability of children's SIP patterns (Conduct Problems Prevention Research Group, 1999; Hudley & Graham, 1993, 1995; Hudley et al., 1998). Most notably, van Schoiack-Edstrom et al. (2002) reported that the Second Step© program resulted in decreases in approval of social exclusion among students in the intervention group relative to controls. Similar results were reported by Nixon and Werner (2009) in an evaluation of a schoolwide relational aggression prevention program. Decreases in sixth graders' beliefs supportive of relational aggression covaried with decreases in relational aggression between pre and posttest, although the lack of control group in this study precludes drawing firm conclusions. The fact that children as young as preschool believe that relational aggression is more acceptable than overt aggression (Goldstein, Tisak, & Boxer, 2002) points to the potential importance of targeting normative beliefs about relational aggression early in development.

As discussed previously, this study has significant conceptual and methodological strengths including the focus on individual and group-level social cognitive processes and the use of independent informants of aggression, and a longitudinal design. It is not without limitations, however. The sample studied here was ethnically homogenous, thus limiting our ability to generalize findings to diverse populations. That being said, little evidence of variation across ethnic groups in relational aggression has been found (e.g., Herrenkohl, Catalano, Hemphill, & Toumbourow, 2009), and Zelli et al. (1999) found no evidence that ethnicity moderated the associations of children's normative beliefs with overt aggression. Nonetheless, these results need replication with a more diverse sample and a longitudinal design that spans multiple years.

Another limitation centers on the assessment of normative beliefs about aggression. The majority of current assessments of SIP patterns with children, including that of normative beliefs, tap into reflective or conscious cognitive processes, despite knowledge that much SIP takes place in an automatic or unconscious manner (e.g., Bargh, 1984; Schneider & Schriffen, 1977). Children's *explicit* evaluations of specific behaviors (assessed in the current study) might differ from their *implicit* representations of relational aggression. It is worth noting that students in this study reported low levels of approval of aggression overall, a pattern that might be explained by our measurement strategy. Clearly, alternative methods for assessing normative beliefs are needed. Burks and colleagues' work on children's internal representations of peers and hostile knowledge structures provides direction for future research (Burks, Dodge, Price, & Laird, 1999; Burks, Laird, Dodge, Pettit, & Bates, 1999). In addition, a recent study with undergraduates used a reading time paradigm and found that students who self-reported higher levels of relational aggression read target sentences depicting aggressive responses to interpersonal conflicts faster than did students reporting lower levels of relational aggression (Linder, Werner, & Lyle, 2009).

Finally, the use of teacher ratings might have underestimated students' actual levels of more covert forms of aggression, particularly among students in middle school. If this is the case, we might expect to find stronger results in future studies using self-reports or peer reports of aggression. On the other hand, if teachers were only able to identify students who openly used relational aggression or who had a pronounced reputation for being relationally aggressive, our results might indicate that peer group norms play an important role for this subgroup of children only, and that the effects reported here would not occur among youth with very low levels of relational aggression. Future research using multiple informants and person-centered analyses to explore possible subgroupings of relational aggression is needed.

References

Bailey, C. A., & Ostrov, J. M. (2008). Differentiating forms and functions of aggression in emerging adults: Associations with hostile attribution biases and normative beliefs. *Journal of Youth and Adolescence, 37*, 713–722.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Bargh, J. A. (1984). Automatic and conscious processing of social information. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 3, pp. 1–43). Hillsdale, NJ: Erlbaum.

Burks, V. S., Dodge, K. A., Price, J. M., & Laird, R. D. (1999). Internal representational models of peers: Implications for the development of problematic behavior. *Developmental Psychology, 35*, 802–810.

Burks, V. S., Laird, R. D., Dodge, K. A., Pettit, G. S., & Bates, J. E. (1999). Knowledge structures, social information processing, and children's aggressive behavior. *Social Development, 8*, 220–236.

Card, N. A., Stucky, B. D., Sawalani, G. M., & Little, T. D. (2008). Direct and indirect aggression during childhood and adolescence: A meta-analytic review of gender differences, intercorrelations, and relations to maladjustment. *Child Development, 79*, 1185–1229.

Cillessen, A. H. N., & Mayeux, L. (2004). From censure to reinforcement: Developmental changes in the association between aggression and social status. *Child Development, 75*, 147–163.

Conduct Problems Prevention Research Group. (1999). Initial impact of the fast track prevention trial for conduct problems: II. Classroom effects. *Journal of Consulting & Clinical Psychology, 67*, 648–657.

Corrigan, A. (2003). *Social Health Profile–Spring, Grade 2/Year 3* (Fast Track Project Technical Report). Retrieved August 15, 2004, from http://www.fasttack-project.org

Crain, M. M., Finch, L. L., & Foster, S. L. (2005). The relevance of the social information processing model for understanding relational aggression in girls. *Merrill-Palmer Quarterly, 51*, 213–249.

Crick, N. R. (1995). Relational aggression: The role of intent attributions, feelings of distress, and provocation type. *Development and Psychopathology, 7*, 313–322.

Crick, N. R. (1996). The role of overt aggression, relational aggression, and prosocial behavior in the prediction of children's future social adjustment. *Child Development, 67*, 2317–2327.

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin, 115*, 74–101.

Crick, N. R., & Grotpeter, J. K. (1995). Relational aggression, gender, and social-psychological adjustment. *Child Development, 66*, 710–722.

Crick, N. R., Grotpeter, J. K., & Bigbee, M. A. (2002). Relationally and physically aggressive children's intent attributions and feelings of distress for relational and instrumental peer provocations. *Child Development, 73*, 1134–1142.

Crick, N. R., & Werner, N. E. (1998). Response decision processes in relational and overt aggression. *Child Development, 69*, 1630–1639.

Delveaux, K. D., & Daniels, T. (2000). Children's social cognitions: Physically and relationally aggressive strategies and children's goals in peer conflict situations. *Merrill-Palmer Quarterly, 46*, 672–682.

Dishion, T. J., Eddy, J. M., Haas, E., Li, F., & Spraklen, K. (1997). Friendships and violent behavior during adolescence. *Social Development, 6*, 207–223.

Dishion, T. J., McCord, J., & Poulin, F. (1999). When interventions harm: Peer groups and problem behavior. *American Psychologist, 54*, 755–764.

Dishion, T. J., Patterson, G. R., & Griesler, P. C. (1994). Peer adaptations in the development of antisocial behavior: A confluence model. In L. R. Huesmann (Ed.), *Aggressive behavior: Current perspectives* (pp. 61–69). New York: Plenum.

Dodge, K. A. (1993). Social-cognitive mechanisms in the development of conduct disorder and depression. *Annual Review of Psychology, 44*, 559–584.

Duffy, A. L., & Nesdale, D. (2008). Peer groups, social identity, and children's bullying behavior. *Social Development, 18*, 121–139.

Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., et al. (1993). Development during adolescence: The impact of stage-environment fit on young adolescents' experiences in schools and in families. *American Psychologist, 48*, 90–101.

Ellis, W. E., & Zarbatany, L. (2007). Peer group status as a moderator of group influence on children's deviant, aggressive, and prosocial behavior. *Child Development, 78*, 1240–1254.

Espelage, D. L., Holt, M. K., & Henkel, R. R. (2003). Examination of peer-group contextual effects on aggression during early adolescence. *Child Development, 74*, 205–220.

Foshee, V. A., Bauman, K. E., Ennett, S. T., Suchindran, C., Benefield, T., & Linder, G. F. (2005). Assessing the effects of the dating violence prevention program "Safe Dates" using random coefficient modeling. *Prevention Science, 6*, 245–258.

Goldstein, S. E., Tisak, M. S., & Boxer, P. (2002). Preschoolers' normative and prescriptive judgments about relational and overt aggression. *Early Education and Development, 13*, 23–39.

Henry, D., Guerra, N., Huesmann, R., Tolan, P., Van Acker, R., & Eron, L. (2000). Normative influences on aggression in urban elementary school classrooms. *American Journal of Community Psychology, 28*, 59–81.

Herrenkohl, T. I., Catalano, R. F., Hemphill, S. A., & Toumbourow, J. W. (2009). Longitudinal examination of physical and relational aggression as precursors to later problem behaviors in adolescents. *Violence and Victims, 24*, 3–19.

Hofmann, D. A., & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management, 24*, 623.

Horton, N. J., & Lipsitz, S. R. (2001). Multiple imputation in practice: Comparison of software packages for regression models with missing variables. *The American Statistician, 55*, 244–254.

Hudley, C., Britsch, B., Wakefield, W. D., Smith, T., Demorat, M., & Cho, S.-J. (1998). An attribution retraining program to reduce aggression in elementary school students. *Psychology in the Schools, 35*, 271–282.

Hudley, C., & Graham, S. (1993). An attributional intervention to reduce peer-directed aggression among African-American boys. *Child Development, 64*, 124–138.

Hudley, C., & Graham, S. (1995). School-based interventions for aggressive African-American boys. *Applied and Preventative Psychology, 4*, 185–195.

Huesmann, R. L. (1988). An information-processing model for the development of aggression. *Aggressive Behavior, 14*, 13–24.

Huesmann, R. L., & Guerra, N. G. (1997). Children's normative beliefs about aggression and aggressive behavior. *Journal of Personality and Social Psychology, 72*, 408–419.

Kuppens, S., Grietens, H., Onghena, P., Michiels, D., & Subramanian, S. V. (2008). Individual and classroom variables associated with relational aggression in elementary-school aged children: A multilevel analysis. *Journal of School Psychology, 46*, 639–660.

Linder, J. R., Werner, N. E., & Lyle, K. A. (2009). *Automatic and controlled social information processing and relational aggression in young adults*. Manuscript submitted for publication.

MacBrayer, E. K., Milich, R., & Hundley, M. (2003). Attributional biases in aggressive children and their mothers. *Journal of Abnormal Psychology, 112*, 698–708.

Murray-Close, D., Crick, N. R., & Galotti, K. M. (2006). Children's moral reasoning regarding physical and relational aggression. *Social Development, 15*, 345–372.

Musher-Eizenman, D. R., Boxer, P., Danner, S., Dubrow, E. F., Goldstein, S. E., & Heretick, D. M. L. (2004). Social-cognitive mediators of the relation of environmental and emotion regulation factors to children's aggression. *Aggressive Behavior, 30*, 389–408.

Muthén, L. K., & Muthén, B. O. (1998–2007). *Mplus user's guide* (5th ed.). Los Angeles, CA: Author.

Nelson, D. A., Mitchell, C., & Yang, C. (2008). Intent attributions and aggression: A study of children and their parents. *Journal of Abnormal Child Psychology, 36*, 793–806.

Nixon, C., & Werner, N. E. (2009). *Reducing adolescents' involvement with relational aggression: Effectiveness of a school-based intervention*. Manuscript submitted for publication.

Pellegrini, A. D., & Bartini, M. (2001). Dominance in early adolescent boys: Affiliative and aggressive dimensions and possible functions. *Merrill-Palmer Quarterly, 47*, 142–163.

Raudenbush, S. W. (1989). Centering predictors in multilevel analysis: Choices and consequences. *Multilevel Modeling Newsletter, 1*(2), 10–12.

Rose, A. J., Swenson, L. P., & Waller, E. M. (2004). Overt and relational aggression and perceived popularity: Developmental differences in concurrent and prospective relations. *Developmental Psychology, 40*, 378–387.

Salmivalli, C., & Voeten, M. (2004). Connections between attitudes, group norms, and behaviour in bullying situations. *International Journal of Behavioral Development, 28*, 3.

Schafer, J. L. (1999). Multiple imputation: A primer. *Statistical Methods in Medical Research, 8*(1), 3.

Schneider, W., & Schriffen, R. M. (1977). Controlled and automatic human information processing: I. Detection, research, and attention. *Psychological Review, 84*, 1–66.

van Schoiack-Edstrom, L., Frey, K. S., & Beland, K. (2002). Changing adolescents' attitudes about relational and physical aggression: An early evaluation of a school-based intervention. *School Psychology Review, 31*, 201–217.

Underwood, M. K., & Buhrmester, D. (2007). Friendship features and social exclusion: An observational study examining gender and social context. *Merrill-Palmer Quarterly, 53*, 412–438.

Werner, N. E., & Crick, N. R. (2004). Maladaptive peer relationships and the development of relational and physical aggression during middle childhood. *Social Development, 13*, 495–514.

Werner, N. E., & Nixon, C. L. (2005). Normative beliefs and relational aggression: An investigation of the cognitive bases of adolescent aggressive behavior. *Journal of Youth and Adolescence, 34*, 229–243.

Werthamer-Larsson, L., Kellam, S., & Wheeler, L. (1991). Effect of first-grade classroom environment on shy behavior, aggressive behavior, and concentration problems. *American Journal of Community Psychology, 19*, 585–602.

Yuan, Y. C. (2000). Multiple imputation for missing data: Concepts and new development. In SAS Institute (Ed.), *Proceedings of the Twenty-Fifth Annual SAS Users Group International Conference*. Cary, NC: SAS Institute.

Zelli, A., Dodge, K. A., Lochman, J. E., & Laird, L. D. (1999). The distinction between beliefs legitimizing aggression and deviant processing of social cues. Testing measurement validity and the hypothesis that biased processing mediates the effects of beliefs on aggression. *Journal of Personality and Social Psychology, 77*, 150–166.

Child Development, May/June 2010, Volume 81, Number 3, Pages 837–850

Antecedents and Correlates of the Popular-Aggressive Phenomenon in Elementary School

Philip C. Rodkin and Glenn I. Roisman
University of Illinois at Urbana-Champaign

This study identified correlates and developmental antecedents that distinguish popular-aggressive elementary school children from other youth. Drawing on the National Institute of Child Health and Human Development Study of Early Child Care and Youth Development ($N = 1022$), popular-aggressive children were identified through teacher ratings over Grades 3–6. Potential correlates included social competence as rated by observers, mothers, and teachers. Potential developmental antecedents included cognitive functioning, maternal sensitivity, and participation in child care through age 4.5. Multinomial regressions allowed for the determination of group differences net of covariates such as gender, ethnicity, maternal education, and income-to-needs. Results showed that popular-aggressive elementary school children were distinguished from other youth as having had an extensive child-care history. Discussion focuses on developmental associations between child-care quantity and aggression.

The paradox of aggressive behavior is that it is simultaneously adaptive and maladaptive (Milgram, 1976). What makes aggression maladaptive is all too obvious: Aggressive children endanger others, and are themselves at risk from early in life for a host of serious adjustment problems. What makes aggression adaptive is also well documented (Rodkin & Wilson, 2007). Aggression is highly evolved over phylogeny (Pellegrini, 2007), is often successful in changing other's behavior (Farmer, Xie, Cairns, & Hutchins, 2007), and can be used to acquire resources and maintain group boundaries (Hawley, 2007). Indeed, the functionality of aggressive behavior is such that Cairns (1979, p. 170) understood the relevant question for behavioral scientists to be not "*why* aggression?" but "why *not* aggression?"

One manifestation of the dual adaptive-maladaptive nature of aggression is the presence of a small group of children who are simultaneously aggressive and popular with their peers (Bagwell, Coie, Terry, & Lochman, 2000; Estell, Cairns, Farmer, & Cairns, 2002; Farmer, Estell, Bishop, O'Neal, & Cairns, 2003; LaFontana & Cillessen, 2002; Lease, Musgrove, & Axelrod, 2002; Luthar & McMahon, 1996; Rodkin, Farmer, Pearl, & Van Acker, 2000). Much of this research has been person oriented, that is, focused on the co-occurrence of high popularity and high aggression among certain children. For example, Rodkin et al. (2000) cluster analyzed teachers' ratings of the popularity, aggression, and other personality characteristics of 450 fourth- through sixth-grade boys, finding popular-aggressive ("tough") and popular-nonaggressive ("model") configurations. Boys in the tough configuration were seen by peers and viewed themselves as popular, cool, aggressive, and athletic—but not prosocial, shy, friendly, or academically oriented. In contrast, model boys were viewed by peers and themselves as popular, athletic, academically oriented, friendly, outgoing, and nonaggressive. Model and tough boys were classroom leaders, affiliating with one another (Farmer et al., 2002) and setting the tone for who and what was cool (Rodkin, Farmer, Pearl, & Van Acker, 2006).

Complementing this literature are variable-oriented studies that show no relation or significant positive associations between continuously distributed measures of popularity and aggression assessed over whole samples (Cillessen & Mayeux,

---

The preparation of this manuscript was supported by grants from the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development to the first (R03 HD048491) and second (U10 HD027040) authors.

Correspondence concerning this article should be addressed to Philip C. Rodkin, 232A Col. Wolfe School, Mail Code 422, 403 E. Healey St., University of Illinois, Champaign, IL 61820, or Glenn I. Roisman, Department of Psychology, University of Illinois, 603 East Daniel Street, Champaign, IL 61820. Electronic mail may be sent to rodkin@illinois.edu or roisman@illinois.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8103-0013

2004; Gest, Sesma, Masten, & Tellegen, 2006; Parkhurst & Hopmeyer, 1998; Prinstein & Cillessen, 2003; Rose, Swenson, & Waller, 2004; Schwartz, Gorman, Nakamoto, & McKay, 2006; Vaillancourt & Hymel, 2006; Xie, Li, Boucher, Hutchins, & Cairns, 2006; Zeller, Vannatta, Schafer, & Noll, 2003). Whether personified in the "tough," popular-aggressive child, or as reflected in bivariate associations, the co-occurrence of popularity and aggression has been demonstrated across early childhood (e.g., Estell et al., 2002), middle childhood (e.g., Rodkin et al., 2000), and adolescence (e.g., Vaillancourt & Hymel, 2006).

Although popular-aggressive children have been identified across development, early factors that enable a coupling between popularity and aggression are not clear (Cillessen & Mayeux, 2007). There are no studies of popular-aggressive children that extend from early childhood to early adolescence, infrequent use of observational measures in this literature (but see Pellegrini, 2007; Pepler, Craig, & Roberts, 1998), and relevant study designs have not assessed broad domains of developmental functioning such as cognition and parenting that at least theoretically might distinguish aggressive children who are versus are not popular. Thus, the goal of this study was to identify correlates and antecedents that distinguish popular-aggressive elementary school children from other youth, particularly other aggressive children who are not popular.

Study questions were addressed using data from the National Institute of Child Health and Human Development (NICHD) Study of Early Child Care and Youth Development (SECCYD). The NICHD SECCYD is well suited for evaluating the early developmental history of popular-aggressive children. Children in the NICHD SECCYD have been followed longitudinally from infancy to age 15 (and age 17.5 year assessment is ongoing). Multiple teacher ratings of popularity and aggression are available. Covariates that may function as selection factors and/or confounds (e.g., sex, ethnicity, family income-to-needs, and maternal education) are included. The sample is also large enough to identify the infrequent configuration of popular-aggressive children. In the present study, the concurrent social competence and disruptiveness of popular-aggressive children as rated by mothers, teachers, and classroom observers was first determined. Two early developmental resources widely regarded as protective of adaptation, high cognitive functioning and sensitive parenting (Masten, 2001), were next evaluated. Finally, effects were determined for the social factor of participation within the early peer ecology of child care, a developmental context that been implicated in the generation of aggressive behavior (Dmitrieva, Steinberg, & Belsky, 2007).

Social competence should correlate with popular aggressiveness. An image of the sophisticated, ruthless, acquisitive child pervades scholarship on the popular-aggressive phenomenon (e.g., Hawley, 2007; Sutton, Smith, & Swettenham, 1999; Vaughn, Vollenweider, Bost, Azria-Evans, & Snider, 2003). Hawley (2002, 2007) identified children who are bistrategic controllers, meaning that they use both prosocial and coercive means to obtain resources. Bistrategic controllers are distinguished from other aggressive children through their savoir faire, ability to impress others, and "moral astuteness" (Hawley, 2007, p. 15). Thus, popular-aggressive children, compared to other aggressive children, should score higher on observer reports of social competence—although perhaps not on measures of social skills as rated by parents and teachers who heavily emphasize compliance.

Developmental antecedents that may distinguish popular-aggressive children from aggressive children who are not popular include early maternal sensitivity and high cognitive functioning. Clearly, the protective resources of high cognitive functioning and maternal sensitivity can promote peer social capital in middle childhood. Studies on popularity and aggression examine factors such as academic engagement and grades (e.g., Rodkin et al., 2000; Schwartz et al., 2006), but for popular-aggressive children, it may be necessary to assess cognition early, before the advent of school disengagement and conflict. The possibility that popular-aggressive children have acquired social resources through effective parenting has simply not been evaluated, as studies of popularity and aggression have tended not to include parenting or family constructs. However, maternal sensitivity has had large and broad effects on developmental adaptation in previous reports of the NICHD Early Child Care Research Network (ECCRN; 2003b; see also Thompson, 2009).

Although it is possible that the popular-aggressive child benefits from broad adaptational resources unavailable to other aggressive children, the main hypothesis that guides this study is that popular-aggressive children will have had more time in child care than other children, including other aggressive children who are not popular. Bronfenbrenner (1979, p. 203) hypothesized that children who spent many hours in child care would be "more likely to engage in egocentric, aggressive, and antisocial behavior both during the preschool years and through later childhood into adolescence.'' Early reports from the NICHD ECCRN confirmed a linkage between child-care quantity and aggression in the SECCYD. When children were in child care at 2 and 3 years of age, child-care quantity (as measured in mean hours per week since birth) was related to negative peer play as rated by caregivers but positive and skilled play as rated by observers (NICHD ECCRN, 2001). When children reached the kindergarten transition at age 4.5, child-care hours were directly associated with externalizing problems, conflicts with adults, and aggression (NICHD ECCRN, 2003a). The association between child-care hours and aggression appears to dissipate by Grade 3 (Belsky et al., 2007; NICHD ECCRN, 2004a, 2005). Nonetheless, some features of child care within the SECCYD continue to show links to aggression into middle childhood (Belsky, 2009). As Bronfenbrenner (1979) predicted, children spending more hours in child care with small- and medium-sized peer groups continue to be rated as more aggressive by Grade 3 teachers (NICHD ECCRN, 2008). Moreover, proportion of time in center-based care has been associated with increased aggression at Grade 6 (Belsky et al., 2007) and lower awakening cortisol levels at age 15 (Roisman et al., 2009) in the SECCYD sample.

The construct of popularity might be able to clarify purported associations between child-care quantity and later aggression. Some children with an extensive child-care background might find that their early expressions of aggression have social value within the child care context and beyond. Dmitrieva et al. (2007), using data from the Early Childhood Longitudinal Study, reported that kindergarteners manifested more externalizing problems by the end of the school year when they had many classmates with extensive child-care histories. This effect of classroom composition was evident even for children who did not themselves attend child care. One interpretation of this finding is that some aggressive children with a history of extensive child care had sufficient popularity and status to influence their peers. Once a linkage between popularity and aggression is established, it may endure over similar peer ecologies with potential for broad social influence.

In this study, NICHD SECCYD participants were classified into popular-aggressive (''tough''), popular-nonaggressive (''model''), aggressive-nonpopular (''aggressive''), and nonpopular-nonaggressive groups using teacher ratings aggregated over Grades 3–6 (see Rodkin et al., 2000, 2006), and then validated through observational assessment. A variety of study measures enabled the determination of the unique developmental effects of early and concurrent cognitive functioning, early and concurrent maternal sensitivity, and early child care (i.e., quantity, quality, center-based care). The robustness of study findings was evaluated by using multiple criteria for popularity and by testing continuous Popularity × Aggression interactions over the entire sample.

Method

Participants

Families were recruited during hospital visits to mothers shortly after the birth of a child in 1991 at 10 locations in the United States. Recruitment and selection procedures are described in detail elsewhere (see http://secc.rti.org; NICHD Early Child Care Research Network, 2003a). The NICHD SECCYD sample is diverse: Twenty-four percent of mothers were minority, 11% of mothers had not completed high school, and 14% were single at the time of the infant's birth. Data for the current report are based on 1,022 children with any data on physically aggressive behavior (as assessed by the Teacher Report Form [TRF]; Achenbach, 1991; ''hits others,'' ''defiant,'' ''mean,'' etc.) and teacher-rated popularity in Grades 3, 4, 5, and 6. These grade school assessments were used due to continuity of measurement (identical measurement of aggression and popularity were available in these grades) and because prior studies of popular aggressiveness have focused on this age range.

Identification of Groups

From the subsample of 1,022 participants, groups were identified based on a cross-classification of teacher-reported aggression and popularity in Grades 3, 4, 5, and 6. The TRF aggression scale showed high internal consistency ($\alpha = .95$ at each assessment). In addition, because TRF reports across Grades 3, 4, 5, and 6 scaled were strongly intercorrelated (mean $r = .54$, range = .50 to .59, all $ps < .001$), a mean TRF aggression score was computed to create groups ($\alpha = .82$). Teachers were asked to rank children's ''popularity with the other same-sex children in your class'' using a measure developed by the NICHD ECCRN and computed at each grade as the sum of the total number of same-sex children plus one, minus study child popularity ranking (teachers ranked all same-sex children in the classroom from most to least popular), all divided by the total number of same-sex children. The range of scores computed in this manner is all positive numbers up to one. Again, because teacher-reported popularity in Grades 3, 4, 5, and 6 were strongly intercorrelated (mean $r = .46$, range = .39 to .52, all $ps < .001$), scores were averaged ($\alpha = .78$) across time to create groups. Note finally that the mean TRF aggression and mean popularity variables were moderately negatively correlated, $r(1{,}022) = -.25$, $p < .001$.

Achenbach (1991) reports that a T score of 60 can be used to define the clinical cut-point of TRF scales. In primary analyses, a cutoff for popularity was used that identified the top third. Using these criteria, four groups were identified as follows: *nonpopular-nonaggressive* ($n = 467$: 240 males, 227 females)—scored below 60 on the TRF and below .66 on teacher-reported popularity; *model* ($n = 418$: 199 males, 219 females)—scored below 60 on the TRF but at or above .66 on teacher-reported popularity; *aggressive* ($n = 104$: 54 males, 50 females)—scored at or above 60 on the TRF but below .66 on teacher-reported popularity; *tough* ($n = 33$: 18 males, 15 females)—scored at or above 60 on the TRF and at or above .66 on teacher-reported popularity. Groups were also recomputed using a more stringent definition of popularity (top 20%; see the Results section); using these more conservative criteria nonpopular-nonaggressive, model, aggressive, and tough groups contained 714, 171, 121, and 16 participants, respectively.

Participants in the analysis subsample ($n = 1{,}022$) differed from those excluded (i.e., the remaining members of the full SECCYD cohort; $N = 1{,}364$) in several ways, although effects were trivial to small. Children in the subsample had higher income-to-needs scores (in sample $M = 3.7$, $SD = 2.8$ vs. excluded $M = 3.2$, $SD = 3.0$), $t(1,\ 1{,}300) = 2.59$, $p < .01$; higher maternal education (in sample $M = 14.5$, $SD = 2.4$ vs. excluded $M = 13.6$, $SD = 2.6$), $t(1,\ 1{,}361) = 5.66$, $p < .001$; were less likely to be male (in sample = 50% vs. excluded = 57%), $\chi^2(1) = 4.64$, $p < .05$; and were less likely to be African American (in sample = 11% vs. excluded = 18%), $\chi^2(1) = 9.88$, $p < .01$.

Procedure

Mothers and children were visited at home and came to university laboratories multiple times from when the children were 1 month to Grade 6. The teachers of participating children completed questionnaires about the children from school entry through Grade 6. The study's Manuals of Operation (http://secc.rti.org) has additional details.

Measures

Demographic characteristics. Child sex and maternal education level were obtained by maternal report at 1 month. Family income was reported by mothers at each major data collection point and converted to an income-to-needs ratio by dividing total family income by the U.S. Census-based poverty-level income for that family size. Ratios were averaged from data obtained at 6–54 months. Child race/ethnicity was coded as 1 = African American, 0 = all others, as previous research suggests that popular-aggressive children may be disproportionately African American (Luthar & McMahon, 1996; Rodkin et al., 2000).

Classroom observations of disruptiveness and social skills. Children's classroom experiences and behavior were observed several times over the course of the NICHD SECCYD. This analysis highlights third- and fifth-grade assessments, which were concurrent with the identification of groups. Using the Classroom Observational Assessment for Third Grade (NICHD ECCRN, 2005) and Fifth Grade (NICHD ECCRN, 2004b), participants were observed for eight 44-min cycles distributed across the school day. Among the 7-point ratings made by observers at each grade were disruptiveness and social/cooperates with peers. In addition to ensuring reliability by having coders rate a set of "gold standard" videos, "live" reliability was ascertained via a second coder. Intercoder intraclass correlations were .87 and .85 for disruptiveness in third and fifth grades, respectively, and .88 and .95 for social/cooperates with peers in third and fifth grades, respectively. Because scores were significantly correlated in third and fifth grades—and analyses conducted separately by assessment yielded an identical pattern of results—two across-time composites were created, standardized on the subsample under investigation: mean disruptiveness, third and fifth grade scores, $r(753) = .18$, $p < .001$, and mean social/cooperates with peers, third and fifth grade scores, $r(928) = .22$, $p < .001$.

Mother and teacher reports of social skills. Mothers completed the 38-item Social Skills Questionnaire from the Social Skills Rating System (SSRS; Gresham & Elliott, 1990) when children were in Grades 3, 4, 5, and 6, indexing general social competence with adults and other children. All items that indexed social competence were summed to obtain a standardized measure of total social skills (αs at the four assessments ranged from .88 to .91), and a composite was created by averaging scores in Grades 3–6 (α = .92). Similarly, teachers completed the 30-item school version of the SSRS (Gresham & Elliott, 1990) when children were in Grades 3, 4, 5, and 6. A standardized total social skills scale was created by summing items indexing social competence (αs at the four assessments ranged from .93 to .94), and a teacher-reported social skills composite was created by averaging scores in Grades 3–6 (α = .77).

Early cognitive functioning. Early cognitive functioning was assessed by standardizing and taking the mean of the total scores on the Bayley Mental Development Index at 15 and 24 months (Bayley, 1991), the Bracken Scale of Basic Concepts (School Readiness Scale; Bracken, 1984) at 36 months, and the Woodcock–Johnson Psycho-Educational Battery–Revised (WJ–R; Woodcock, 1990) at 54 months (α = .79).

Later cognitive functioning. Cognitive functioning in grade school was assessed by standardizing and taking the mean of WJ–R (Woodcock, 1990) scores administered to participants in Grades 1, 3, and 5. W (standard) scores for subscales at each time point were standardized and averaged (subscale reliability within time ranged from .84 to .94 in the subsample) before compositing across the three measurement occasions (α = .95).

Early maternal sensitivity. Early maternal sensitivity was calculated from ratings of mothers' behavior during a videotaped interaction between mother and child under semi-structured, free-play conditions at 6, 15, 24, 36, and 54 months (detailed in NICHD Early Child Care Research Network, 1998, 2002, 2003a, 2003b). The early maternal sensitivity score represents child-centered interactions that: reflect emotional support for the child that is responsive to the child's needs and warmly encouraging of the child's positive engagement with the toys and tasks in the interaction procedures (sensitivity to nondistress, supportive presence), is respectful of the child's interests and efforts (lack of intrusiveness, respect for autonomy), and is affectively positive (positive regard, lack of hostility). Internal consistency of the sensitivity composite ranged from αs = .70 to .84, and interrater reliabilities of the composites, calculated with intraclass correlations across raters, ranged from .83 to .87. Mean (standardized) sensitivity scores at 6, 15, 24, 36, and 54 months were used in analyses as a measure of early sensitivity (α = .76).

Later maternal sensitivity. In Grade 1, NICHD ECCRN staff videotaped an interaction between mother and child under semi-structured, free-play conditions. In Grades 3 and 5, the context shifted to two tasks requiring participants to discuss areas of disagreement and to engage in a planning exercise. Sensitivity was computed as a composite of supportive presence, respect for autonomy, and hostility (reversed) at each assessment. Internal consistency of the sensitivity composite ranged from αs = .80 to .85, and interrater reliabilities of the composites, calculated with intraclass correlations across raters, ranged from .84 to .91. Mean (standardized) sensitivity scores in Grades 1, 3, and 5 were used in analyses as a measure of later sensitivity (α = .71).

Early child care. Child-care information was obtained through phone calls with the mother and observations of the child's primary child-care setting. Mothers were called every 3 months between the time the baby was 1 month and 36 months, and at 4-month intervals through 54 months, and asked to list the various places the child received care and the hours per week that the child spent in each arrangement. The quantity of care is the average hours per week from 1 to 54 months that the child spent in regular nonmaternal care; the amount of center care is expressed as the proportion of phone calls from 1 to 54 months in which the child was reported to be attending a child-care center for at least 10 hr/week.

Quality of care was assessed via live observation of primary nonmaternal arrangements that were used for 10 or more hr per week at 6, 15, 24, 36, and 54 months, focusing on caregiver behaviors toward the study child. Observations were conducted during two half-day visits scheduled within a 2-week interval at 6, 15, 24, and 36 months and one half-day visit at 54 months. At each half-day visit, observers completed two 44-min cycles of the Observational Record of the Caregiving Environment (ORCE; NICHD ECCRN, 1996, 2002). Reliability exceeded .90 at 6 months, .86 at 15 months, .81 at 24 months, .80 at 36 months, and .90 at 54 months. The aggregated scores from 6, 15, 24, 36, and 54 months were standardized and averaged to form the child-care quality composite.

Results

Analytic Plan

Planned comparisons, which are ideal in protecting against Type II error when a priori contrasts can be specified (e.g., they eliminate the need for omnibus *F*s; see Hays, 1988), were conducted in preliminary analyses. Each set of analyses contrasts the popular-aggressive (i.e., "tough") group with the other three groups (e.g., nonpopular-nonaggressive, model, and aggressive). In addition to *t* statistics and their associated two-tailed *p* values, standardized differences between means (Cohen's *d*; effect sizes) were calculated. Note that (a) *n*s vary across analysis to maximize data available from each assessment (see Table 1, which presents *n*s, *M*s, *SD*s for all dependent measures by group) and (b) because Levene's test of homogeneity was significant for all planned comparisons except observed social skills, child-care quantity, and child-care quality, *t*s reported for all other planned comparisons do not assume equal variances across groups and *df*s are estimates rounded to the nearest whole number.

Analyses are reported in five sets. First, the tough group was externally validated by examining convergent evidence drawn from contemporaneous measures from Grades 3 and 5. The question here was whether children in the tough group are in fact simultaneously aggressive and socially effective—although not necessarily socially skilled based on adult standards. The second set of analyses examined whether two crucial developmental resources—adequate or better antecedent and contemporaneous parenting and cognitive functioning—served as explanatory factors in accounting for the presence of tough children. The third set of analyses examined early child-care experiences as an alternative explanation for tough children's co-occurring aggression and popularity.

The fourth set of analyses turns to the unique effects of predictors in aggregate using multinomial regression analysis with the tough participants as the reference group. These analyses were conducted to examine which of the developmental resources (i.e., contemporaneous and antecedent cognitive functioning and sensitivity) and alternative explanatory factors (i.e., child-care experiences) accounted for popular-aggressive status in grade school, even after controlling for background characteristics that might function as selection factors or confounds (i.e., demographic covariates; child sex, child African American vs. not, maternal education, and income-to-needs from 6 to 54 months). Fifth, to confirm that findings can not be attributed to the specific ways in which groups were created, robustness analyses considered whether: (a) groups constructed using a more conservative threshold for defining popularity (i.e., top 80% of class as compared to top third) yielded convergent evidence in follow-up multinomial regressions and (b) multinomial regression results could be replicated in a variable-based regression examining the predictive significance of the interaction of popularity and

Table 1
Dependent Measures by Group

	Nonpopular-nonaggressive		Model		Aggressive		Tough	
Variable	*M* (*SD*)	*n*	*M* (*SD*)	*n*	*M* (*SD*)	*n*	*M* (*SD*)	*n*
Concurrent indicators								
Obs. disruptiveness	−0.09 (0.87)	457	−0.13 (0.75)	413	+0.65 (1.45)	100	+0.96 (2.06)	33
Obs. social skills	−0.21 (0.97)	456	+0.28 (0.95)	410	−0.26 (1.03)	100	+0.21 (1.00)	33
T social skills	100 (9.6)	467	110 (8.1)	418	87 (8.8)	104	91 (8.9)	33
M social skills	106 (14.3)	463	111 (12.4)	413	96 (14.9)	103	100 (15.9)	31
Developmental resources								
Early cognition	−0.09 (1.00)	459	+0.24 (0.90)	413	−0.48 (1.12)	100	−0.27 (0.99)	32
Later cognition	−0.05 (0.96)	464	+0.20 (0.89)	415	−0.45 (1.27)	104	−0.34 (1.29)	32
Early sensitivity	+0.01 (0.96)	464	+0.19 (0.92)	417	−0.63 (1.15)	102	−0.54 (1.15)	33
Later sensitivity	−0.00 (1.00)	463	+0.19 (0.88)	413	−0.56 (1.12)	102	−0.62 (1.19)	31
Early child-care experiences								
Hr/week in CC	23.5 (14.0)	467	23.9 (14.1)	418	23.0 (12.9)	104	31.8 (12.8)	33
Quality of CC	−0.02 (1.01)	423	+0.11 (0.98)	374	−0.27 (1.01)	93	−0.27 (0.88)	33
% CC	0.20 (0.24)	451	0.21 (0.27)	411	0.23 (0.24)	99	0.30 (0.29)	32

Note. Mean values reported for all observational measures are *z* scores standardized on the subsample classified into one of the groups. Groups were defined using a 66% cutoff for popularity and $T \geq 60$ cutoff for defining significantly elevated levels of aggression. Obs = observed; T = teacher; M = mother; CC = child care.

aggressiveness (controlling for main effects) over the entire analysis sample.

Contemporaneous Correlates

Observations of classroom behavior. Observers rated tough children as significantly more disruptive during Grades 3 and 5 than model children, $t(33) = 3.02$, $p < .01$, $d = 0.70$, and nonpopular-nonaggressive children, $t(33) = 2.91$, $p < .01$, $d = 0.66$, but not more disruptive than aggressive children, $t(43) = 0.79$, $p = .43$, $d = 0.17$. Consistent with their status as popular, observers rated tough children as significantly *more* socially skilled than nonpopular-nonaggressive, $t(995) = 2.40$, $p < .05$, $d = 0.43$, and aggressive children, $t(995) = 2.43$, $p < .05$, $d = 0.47$, and comparable to model children, $t(995) = 0.38$, $p = .70$; $d = 0.07$.

Teacher and mother ratings of social skills. Teachers rated tough children as significantly less socially skilled on the SSRS in Grades 3–6 than model children, $t(36) = 11.89$, $p < .001$, $d = 2.24$, and nonpopular-nonaggressive children, $t(38) = 5.70$, $p < .001$, $d = 0.99$, although *more* socially skilled than aggressive children, $t(53) = 2.58$, $p < .05$, $d = 0.52$. Mothers rated tough children as significantly less socially skilled in Grades 3–6 than model children, $t(33) = 3.79$, $p < .001$, $d = 0.78$, and nonpopular-nonaggressive children, $t(33) = 2.34$, $p < .05$, $d = 0.45$, and comparable to aggressive children, $t(47) = 1.24$, $p = .22$; $d = 0.26$.

Developmental Resources

Cognitive functioning. Prior to grade school, tough children had significantly lower levels of cognitive functioning than model children, $t(35) = 2.84$, $p < .01$, $d = 0.54$, but were comparable to nonpopular-nonaggressive children, $t(36) = 1.00$, $p = .32$, $d = 0.18$, and aggressive children, $t(59) = 1.01$, $p = .32$, $d = 0.20$. Results in grade school were identical. Tough children had significantly lower levels of cognitive functioning than model children, $t(33) = 2.31$, $p < .05$, $d = 0.49$, but were comparable to nonpopular-nonaggressive children, $t(33) = 1.25$, $p = .22$, $d = 0.26$, and aggressive children, $t(51) = 0.41$, $p = .69$, $d = 0.08$.

Maternal sensitivity. Prior to grade school, tough children had significantly less sensitive mothers than model children, $t(35) = 3.53$, $p < .01$, $d = 0.70$, and nonpopular-nonaggressive children, $t(35) = .68$, $p < .05$, $d = 0.52$, but were comparable to aggressive children, $t(54) = 0.38$, $p = .70$, $d = 0.08$. Results in grade school were identical. Tough children had significantly less sensitive mothers than model children, $t(33) = 3.69$, $p < .001$, $d = 0.77$, and nonpopular-nonaggressive children, $t(33) = 2.81$, $p < .01$, $d = 0.56$, but were comparable to aggressive children, $t(47) = 0.24$, $p = .81$, $d = 0.05$.

Early Child-Care Experiences

Groups were examined on three measures of child care measured through 54 months: hours per week in nonmaternal child care (quantity), quality, and proportion of time in center-based care. Tough children spent significantly more time in nonmaternal child care than model, $t(1018) = 3.12$, $p < .01$, $d = 0.58$; nonpopular-nonaggressive, $t(1018) = 3.30$, $p < .001$, $d = 0.62$, and aggressive, $t(1018) = 3.16$, $p < .01$, $d = 0.68$, children. Tough children experienced significantly lower quality child care than model children, $t(919) = 2.13$, $p < .05$, $d = 0.41$, although the quality of their child care was comparable to nonpopular-nonaggressive children, $t(919) = 1.39$, $p = .16$, $d = 0.26$, and aggressive children, $t(919) = 0.02$, $p = .99$, $d = 0.004$. Tough children spent a marginally larger proportion of time in center-based care than model children, $t(35) = 1.70$, $p = .098$, $d = 0.32$, and nonpopular-nonaggressive children, $t(34) = 1.98$, $p = .06$, $d = 0.39$, but comparable time as aggressive children, $t(45) = 1.24$, $p = .22$, $d = 0.26$.

Aggregate Effects of Predictor Variables

A multinomial regression predicting group was conducted with the tough group as the reference category (see Table 2). We simultaneously entered four demographic covariates (child sex, maternal education, family income-to-needs, African American vs. not), the entire set of developmental resource variables (early and later cognitive functioning and maternal sensitivity) and the early child-care variables (quantity, quality, and proportion of center-based care). The overall model fit the data well, $\chi^2(33, N = 898) = 127.36$, $p < .001$, and the Nagelkerke pseudo-R^2 was .15 (details regarding likelihood ratio tests are available from the authors). Three sets of parameter estimates are provided in Table 2 (nonpopular-nonaggressive vs. tough; model vs. tough; and aggressive vs. tough). As can be seen, only two variables consistently distinguished the tough children from each of the other groups. Specifically, net of other effects, tough children were disproportionately African American and spent more hours per week in nonmaternal child care before they transitioned to grade school (i.e., through 54 months). In addition, a follow-up analysis examined whether the significant effect of child-care quantity was moderated either by child-care quality or center-based care exposure in a simplified multinomial regression in which the three main effects of early child care (quantity, quality, and center care) and the two relevant interaction effects (Quantity × Quality and Quantity × Center-Based Care) were entered. Neither interaction effect was significant, but the main effect of quantity continued to distinguish tough children from other groups.

Robustness Analyses

To examine whether study findings were a product of the top 33% popularity criterion used in primary analyses, the multinomial regression of Table 2 was recomputed using a more stringent criterion that required "popular" participants to be in the top 20% in teacher-related popularity. As detailed in Table 3, the pattern of results was largely unchanged, suggesting that results are largely insensitive to the popularity criterion. Specifically, the model fit the data well, $\chi^2(33, N = 898) = 116.10$, $p < .001$, and the Nagelkerke pseudo-R^2 for the model was .15 (details regarding likelihood ratio tests are available from the authors). Once again, only percent African American and early child-care quantity consistently discriminated the tough children from all other groups. Of note, however, African Americans were only marginally overrepresented in the tough versus aggressive group in this analysis.

Two multiple regressions examined whether early child care and percent African American were associated with simultaneously high aggression and popularity levels in variable-based analyses. First, child-care quantity was regressed onto mean aggression (Grades 3–6), mean popularity (Grades 3–6), and their interaction ($R^2 = .01$, $p < .01$). Results showed significant main effects of aggression ($\beta = .11$, $p < .01$) and popularity ($\beta = .07$, $p < .05$), as well as the expected significant interaction between popularity and aggression ($\beta = .07$, $p < .05$). The interaction is graphed in Figure 1 along Aiken and West's (1991) guidelines (i.e., ± 1 *SD*) and indicates that higher aggression and more child care are correlated only for more popular children. In contrast, a binary logistic regression examining whether the interaction of popularity and aggression was correlated with percent African American (Nagelkerke $R^2 = .18$, $p < .001$) yielded evidence only for a sig-

Table 2

Parameter Estimates for Multinomial Logistic Regression Predicting Popular-Aggressive Groups

Variable	*B*	*SE*	Wald	*df*	*p*	e^B
Nonpopular-nonaggressive vs. tough						
Intercept	3.03	1.67	3.29	1	.07	
Sex	0.21	0.41	0.26	1	.61	1.23
Maternal education	−0.06	0.11	0.24	1	.62	0.95
Income-to-needs	0.06	0.12	0.23	1	.63	1.06
African American?	−2.43	0.53	20.69	1	< .001	0.09
Early cognitive Functioning	−0.52	0.31	2.86	1	.091	0.60
Later cognitive Functioning	0.10	0.28	0.14	1	.71	1.11
Early sensitivity	0.25	0.26	0.91	1	.34	1.28
Later sensitivity	0.22	0.23	0.92	1	.34	1.24
Hr/week in CC	−0.05	0.02	5.49	1	< .05	0.96
Quality of CC	−0.11	0.21	0.26	1	.61	0.90
% CC	−0.71	0.75	0.88	1	.35	0.49
Model vs. tough						
Intercept	4.08	1.67	5.97	1	< .05	
Sex	0.24	0.41	0.34	1	.56	1.27
Maternal education	−0.11	0.12	0.90	1	.34	0.90
Income-to-needs	0.09	0.12	0.60	1	.44	1.09
African American?	−1.69	0.53	10.06	1	< .01	0.18
Early cognitive Functioning	−0.16	0.31	0.25	1	.61	0.86
Later cognitive Functioning	0.23	0.28	0.68	1	.41	1.26
Early sensitivity	0.26	0.26	1.01	1	.32	1.30
Later sensitivity	0.31	0.23	1.86	1	.17	1.37
Hr/week in CC	−0.04	0.02	4.94	1	< .05	0.96
Quality of CC	−0.06	0.22	0.07	1	.80	0.95
% CC	−0.62	0.76	0.68	1	.41	0.54
Aggressive vs. tough						
Intercept	2.56	1.82	1.99	1	.16	
Sex	0.17	0.45	0.15	1	.70	1.19
Maternal education	−0.09	0.13	0.49	1	.49	0.92
Income-to-needs	0.02	0.13	0.02	1	.89	1.02
African American?	−1.65	0.58	8.21	1	< .01	0.19
Early cognitive Functioning	−0.39	0.34	1.33	1	.25	0.68
Later cognitive Functioning	0.04	0.30	0.02	1	.88	1.05
Early sensitivity	−0.02	0.28	0.01	1	.95	0.98
Later sensitivity	−0.01	0.25	0.00	1	.97	0.99
Hr/week in CC	−0.05	0.02	6.08	1	< .05	0.95
Quality of CC	−0.15	0.23	0.41	1	.52	0.84
% CC	0.06	0.84	0.01	1	.95	1.06

Note. For this analysis, $n = 30$ for tough, $n = 414$ for nonpopular-nonaggressive, $n = 365$ for model, and $n = 89$ for aggressive. Groups were defined using a 66% cutoff for popularity. Positive *B*s indicate that the group scored higher on the variable in question compared to the reference group (popular-aggressive) net of other variables. For sex, 1 = male, 0 = female. For African American, 1 = African American, 0 = not African American, CC = child care.

nificant main effect of aggression ($e^B = 2.42$, $p < .001$; African American children were viewed by teachers as more aggressive), but not a main effect of popularity ($e^B = 1.17$, $p = .19$) or a significant Aggression × Popularity interaction ($e^B = 1.16$, $p = .11$).

Table 3

Parameter Estimates for Multinomial Logistic Regression Predicting Popular-Aggressive Groups

Variable	*B*	*SE*	Wald	*df*	*p*	e^B
Nonpopular-nonaggressive vs. tough						
Intercept	4.33	2.42	3.20	1	.07	
Sex	0.18	0.59	0.11	1	.76	1.20
Maternal education	0.01	0.16	0.00	1	.97	1.01
Income-to-needs	−0.05	0.14	0.16	1	.69	0.95
African American?	−2.30	0.77	8.85	1	< .01	0.10
Early cognitive Functioning	−0.73	0.45	2.62	1	.11	0.48
Later cognitive Functioning	−0.48	0.42	1.32	1	.25	0.62
Early sensitivity	0.72	0.37	3.85	1	.05	2.06
Later sensitivity	0.03	0.34	0.01	1	.92	1.03
Hr/week in CC	−0.07	0.03	5.88	1	< .05	0.93
Quality of CC	0.10	0.30	0.11	1	.75	1.10
% CC	1.38	1.22	1.27	1	.26	3.97
Model vs. tough						
Intercept	3.76	2.50	2.27	1	.13	
Sex	0.06	0.61	0.01	1	.92	1.06
Maternal education	−0.06	0.16	0.13	1	.72	0.94
Income-to-needs	−0.01	0.14	0.01	1	.92	0.99
African American?	−1.99	0.84	5.65	1	< .05	0.14
Early cognitive Functioning	−0.44	0.46	0.89	1	.34	0.65
Later cognitive Functioning	−0.44	0.43	1.05	1	.31	0.64
Early sensitivity	0.86	0.39	4.98	1	< .05	2.37
Later sensitivity	0.14	0.35	0.15	1	.70	1.15
Hr/week in CC	−0.07	0.03	5.03	1	< .05	0.94
Quality of CC	0.21	0.31	0.47	1	.49	1.24
% CC	1.12	1.26	0.79	1	.38	3.05
Aggressive vs. tough						
Intercept	3.10	2.53	1.50	1	.22	
Sex	0.10	0.62	0.03	1	.88	1.10
Maternal education	0.01	0.17	0.00	1	.97	1.01
Income-to-needs	−0.12	0.15	0.65	1	.42	0.89
African American?	−1.51	0.80	3.53	1	.06	0.22
Early cognitive Functioning	−0.69	0.47	2.14	1	.14	0.50
Later cognitive Functioning	−0.65	0.43	2.32	1	.14	0.52
Early sensitivity	0.54	0.39	1.99	1	.16	1.72
Later sensitivity	−0.25	0.35	0.51	1	.48	0.78
Hr/week in CC	−0.07	0.03	6.08	1	< .05	0.93
Quality of CC	0.09	0.31	0.08	1	.78	1.09
% CC	2.34	1.28	3.35	1	.07	10.39

Note. For this analysis, $n = 14$ for tough, $n = 627$ for nonpopular-nonaggressive, $n = 152$ for model, and $n = 105$ for aggressive. Groups were defined using an 80% cutoff for popularity. Positive *B*s indicate that the group scored higher on the variable in question compared to the reference group (popular-aggressive) net of other variables. For sex, 1 = male, 0 = female. For African American, 1 = African American, 0 = not African American, CC = child care.

Discussion

This study was conducted in order to learn more about a small but influential group of children who paradoxically are both very aggressive and very popular with their peers. The first study question was whether the social competencies that popular-aggressive ("tough") children enjoy are visible to objective observers, and to the mothers and teachers of these children. The second question we explored was whether the developmental resources of high cognitive functioning and maternal sensitivity distinguish tough children from the majority of aggressive children who are not popular. The third question, and the primary focus of this study, was whether tough children might have had a unique history of extensive participation in early child care through the age of 54 months. Research questions were addressed by making use of the public access data of the NICHD SECCYD through Grade 6. The NICHD SECCYD allowed study analyses to take account of many features that are not typical in studies of the popularity-aggression phenomenon, such as longitudinal data at regular intervals from infancy to Grade 6, measures of popularity and aggression aggregated over four annual assessments from Grades 3 to 6, classroom observations aggregated over Grades 3 and 5, and information on potential covariates and multiple domains of

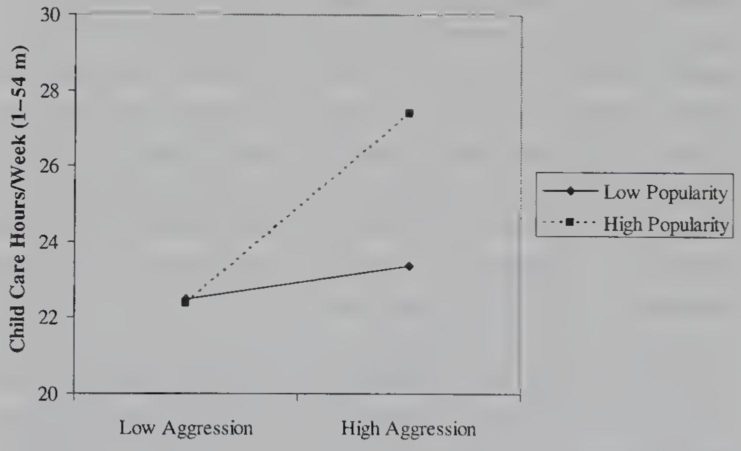


Figure 1. Associations between early child-care quantity (1–54 months) and the interaction of popularity and aggression (Grades 3–6), computed over the entire analysis sample using Aiken and West's (1991) guidelines (i.e., "High" = +1 *SD*, "Low" = −1 *SD*).

developmental functioning. Study findings indicated one correlate that consistently distinguished tough children from all other children: an extensive early history of nonmaternal child care.

The first study question addressed validity by asking whether children rated as popular and aggressive by teachers showed additional evidence of social competence. Can observers discern the social engagement that peer popularity entails, even when popularity is conjoined with aggression? Indeed, they can. Tough children were observed to be extremely disruptive yet highly sociable and cooperative with their peers. According to observers, tough and ''model'' (popular-nonaggressive) children had similar and high levels of social competence—and higher levels of social competence than nonpopular children. Do teachers and mothers recognize the social competencies of the popular-aggressive children in their midst? No, they do not. Mothers and teachers rated tough children as having poorer social skills than model and nonpopular-nonaggressive children. Thus, tough children were viewed as socially skilled and aggressive by disinterested observers who focused on peer interaction quality, but as not socially skilled by adults who may have a frustrating history of trying to guide tough children toward prosocial behavior.

The second study question asked whether tough children enjoy developmental resources similar to model children, and more resources than aggressive children who are not popular. Given the interpersonal skills of tough children, might they have a history of high cognitive functioning? Might tough children benefit from effective, sensitive parenting, in contrast to other aggressive children who suffer from hostile, inconsistent parenting? To both these questions, the answer appears to be no. Model children had higher levels of cognitive functioning and more sensitive parenting than tough children, who were similar to aggressive-nonpopular children in these respects. Multinomial regressions did not show earlier or later cognitive functioning or maternal sensitivity to uniquely distinguish tough children. Aggressive children had relatively low cognitive functioning and insensitive parenting irrespective of their popularity.

The third study question examined three features of children's early child care: quantity, quality, and proportion of center-based care. Children who were popular and aggressive in middle childhood had an extensive history of early child care. From the period of 1–54 months of age, popular-aggressive children had an average of 31.8 hr per week of child care compared to 23.6 hr for other children. Using a more stringent criterion for popularity (top 20% vs. top third), the effect of child-care quantity increases to an average of 34.6 hr per week for popular-aggressive children and 23.7 hr per week for all other children, a difference of almost 11 hr per week. Multinomial regressions confirmed child-care quantity differences to be significant when Tough children were contrasted with aggressive-nonpopular, popular-nonaggressive, and nonpopular-nonaggressive children net of other factors. Moreover, when popularity and aggression were treated continuously over the entire sample, there was a significant Popularity × Aggression interaction, such that more hours of early child care were associated with the combination of high popularity and high aggression in middle childhood.

Why might an extensive history of early child care distinguish popular-aggressive children from other children, including other aggressive children, over the grade school years? Three factors may be considered: the sometimes aggressogenic nature of early child-care contexts, the relation between the prevalence of aggression and its peer acceptance, and the enduring effects of aggression that has an actual or perceived base of social support.

Early childhood peer ecologies can sometimes appear to be conducive to the development of aggressive behavior. Children who spend a large part of their day in child care may be in the best position to learn how to employ aggression so as to garner social and material resources. Here, it is worth keeping in mind Luthar and McMahon's (1996, p. 601) conclusion that ''what is valued or admired within the peer group is often diametrically opposed to what is viewed as positive adjustment from the standpoint of adults in conventional mainstream society.'' Luthar and McMahon were speaking of adolescents, but their comments hold true for early childhood as well. Aggressive, agonistic behavior can promote popularity, dominance, and status throughout the early childhood years (Estell et al., 2002; Hawley, 2002; Murphy, Murphy, & Newcomb, 1937; Pellegrini et al., 2007; Stormshak, Bierman, Bruschi, Dodge, Coie, & Conduct Problems Prevention Research Group, 1999; Strayer, 1980; Vaughn et al., 2003).

There is reason to suspect that the early peer ecologies of SECCYD participants may sometimes have been aggressogenic, so that children with extensive child care may have had greater potential to conjoin aggressive behavior with peer (but not adult) perceived social competence. For example, NICHD Early Child Care Research Network (2001) focused on relations between time in child care with other children and peer interaction skills at 2 and 3 years of age. Average weekly hours of care since birth was associated with negative peer play as rated by caregivers (e.g., teases, bullies others) and positive, skilled play as rated by observers (e.g., complex, reciprocal, frequent engagement). NICHD Early Child Care Research Network questioned these results: "How could caregivers perceive and rate children as more negative, whereas observers rated the same children's actual interactions in the same setting as more positive?" The answer may be that children who spend well over 30 hr per week in child care have had time and opportunity to master a peer reward structure in which aggressors can become socially well integrated, but nonetheless problematic for caregivers concerned with order and positive socialization.

The developmental question of whether child care–aggression linkages dissipate, endure, or spread might best be answered by reconsidering Bronfenbrenner's (1979) hypothesis that the relation between child-care quantity and later aggression is elicited and continually supported by peer-based social experiences (see also Belsky, 2009; Belsky et al., 2007; Dmitrieva et al., 2007; Fabes, Hanish, & Martin, 2003; Maccoby & Lewis, 2003). Although the main effect between child-care quantity and later aggression in the SECCYD dissipates after kindergarten (NICHD Early Child Care Research Network, 2004a, 2005), it could be that the effect becomes concentrated toward child-care arrangements with peers, such as: nonrelative care (NICHD Early Child Care Research Network, 2005), center-based care (Belsky et al., 2007; NICHD Early Child Care Research Network, 2005), and quantity of time in child care with small- and medium-sized peer groups (NICHD Early Child Care Research Network, 2008). The present study suggests that a carrier of these peer group effects may be the popular-aggressive child, who has spent on average 8–11 more hours a week in child care than other children over a 4½-year period. By Grade 6, these tough children may have enduring patterns of peer aggression, possibly influencing other children toward antisocial behavior as well.

Of course, the mechanisms proposed here are speculative. Moreover, in some cases the child care–aggression association is null, or negative. For example, Côté et al. (2007) reported that nonmaternal child care *reduced* the risk of later aggression when care commenced before 9 months of age, but only for children with mothers with less than a high school education. Additionally, positive effects of high quality child care have been demonstrated within the SECCYD and other samples, including more advanced linguistic functioning and higher mathematics and reading achievement (e.g., Dmitrieva et al., 2007; NICHD Early Child Care Research Network, 2005).

It is important also to emphasize that the effects found in this report are not exhaustive of all possible developmental antecedents of popular aggressiveness. For example, Vaillancourt and Hymel (2006) suggest that popularity is associated not only with aggression (i.e., "explicit power") but also with peer-valued characteristics (i.e., "implicit power") such as being stylish, good looking, athletic, funny, tough, and wealthy (see also Adler & Adler, 1998; Sippola, Paget, & Buchanan, 2007). Athleticism, attractiveness, material resources, and related qualities are clearly important from early childhood to adolescence and beyond (Hawley, Johnson, Mize, & McNamara, 2007; Pellegrini, 2007; Vaughn et al., 2003), but the NICHD SECCYD does not include many measures of the developmental resource of peer-valued characteristics.

Additionally, aggression and popularity measures drawn from the NICHD SECCYD were reliable, but were assessed in ways that are not typical in the peer relations literature. The focus of this report was on physical aggression as represented by the Achenbach (1991) TRF aggression scale. Along with not including many items on relational, social, or indirect aggression, the TRF aggression scale has limited capacity to detect gender differences because the scale is standardized by gender. Popularity was rated by teachers, as the NICHD SECCYD does not include peer nominations in middle childhood. Peer nominations represent a gold standard in the peer relations literature (but see Olweus, 2010), but previous studies on popular aggressiveness have used teacher ratings (e.g., Rodkin et al., 2000), and there is generally strong concordance between teacher and peer representations of classroom social networks (Gest, 2006; Pearl, Leung, Van Acker, Farmer, & Rodkin, 2007). On a related note, a clear limitation of the present analysis is that the NICHD SECCYD does not include extensive measures of the peer social networks of early child-care environments. Without these data, it remains speculative that children who are popular and aggressive in middle childhood are also central in early childhood peer ecologies (see also Lintong & van IJzendoorn, 2009). Dynamic peer network data throughout the child care and grade school years would yield rich information in which social influence patterns could be developmentally mapped and modeled (Hanish & Rodkin, 2007).

Bronfenbrenner (1944, 1979), an early scholar of social status, warned that aggression could emerge as an adaptive short-term response for young children with extensive child-care experiences (see also Belsky & Steinberg, 1978). The aggression discussed in pre-NICHD SECCYD reports of child-care quantity corresponds with the kind of instrumental, proactive aggression typically associated with peer status—social skills (Haskins, 1985), ''fearless aggressive nonconformity'' (Moore, 1975, p. 257), and ''independence rather than disturbed behavior'' (Clarke-Stewart, 1989, p. 269). The aggression-as-adaptation perspective lost ground to deficit approaches once child-care quantity failed to differentially predict externalizing subscales of the Achenbach (1991) at age 4.5 that theoretically ranged from ''more adaptive'' and less risky (i.e., assertiveness) to ''less adaptive'' and more risky (i.e., aggression, the same TRF subscale used in the present study over middle childhood; NICHD Early Child Care Research Network, 2003a, Table 9).

However, the adaptational significance of childhood aggression depends less on its form than on *who* is aggressing and the *context* in which aggression is elicited. This restatement of Wright, Giammarino, and Parad's (1986) person-group similarity model (see also Chang, 2004; Stormshak et al., 1999) may hold true for many educational settings in America. Bronfenbrenner (1979) and Maccoby and Lewis (2003) stressed that quality child care could support socioemotional development under certain conditions, including knowledgeable scaffolding for the peer ecologies of youth. Indeed, the prevalence of the popular-aggressive child reminds us of the challenge that still awaits: To design educational systems in which children and adults share the same goals of learning and character, a challenge first pointed out by Coleman (1961) almost 50 years ago. As for the popular-aggressive children of the NICHD SECCYD, time will tell whether their social status can be transformed into social capital once high school is left behind, and the peer ecology of the school is replaced by new social challenges in emerging adulthood.

References

Achenbach, T. M. (1991). *Integrative guide for the 1991 CBCL/4-18, YSR, and TRF Profiles*. Burlington: University of Vermont, Department of Psychiatry.

Adler, P. A., & Adler, P. (1998). *Peer power: Preadolescent culture and identity*. New Brunswick, NJ: Rutgers University Press.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Bagwell, C. L., Coie, J. D., Terry, R. A., & Lochman, J. E. (2000). Peer clique participation and social status in preadolescence. *Merrill-Palmer Quarterly*, *46*, 280–305.

Bayley, N. (1991). *Bayley Scales of Infant Development: Second Edition (Standardization Version)*. New York: Psychological Corporation.

Belsky, J. (2009). Classroom composition, childcare history and social development: Are childcare effects disappearing or spreading? *Social Development*, *18*, 230–238.

Belsky, J., & Steinberg, L. D. (1978). The effects of day care: A critical review. *Child Development*, *49*, 929–949.

Belsky, J., Vandell, D. L., Burchinal, M., Clarke-Stewart, K. A., McCartney, K., Owen, M. T., et al. (2007). Are there long-term effects of early child care? *Child Development*, *78*, 681–701.

Bracken, B. A. (1984). *Bracken Basic Concept Scale*. San Antonio, TX: Psychological Corporation.

Bronfenbrenner, U. (1944). A constant frame of reference for sociometric research: II. Experiment and inference. *Sociometry*, *7*, 40–75.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.

Cairns, R. B. (1979). *Social development: The origins and plasticity of interchanges*. San Francisco: W.H. Freeman.

Chang, L. (2004). The role of classroom norms in contextualizing the relations of children's social behaviors to peer acceptance. *Developmental Psychology*, *40*, 691–702.

Cillessen, A. H. N., & Mayeux, L. (2004). From censure to reinforcement: Developmental changes in the association between aggression and social status. *Child Development*, *75*, 147–163.

Cillessen, A. H. N., & Mayeux, L. (2007). Variations in the association between aggression and social status: Theoretical and empirical perspectives. In P. Hawley, T. D. Little, & P. C. Rodkin (Eds.), *Aggression and adaptation: The bright side to bad behavior* (pp. 135–156). Mahwah, NJ: Erlbaum.

Clarke-Stewart, K. A. (1989). Infant day care: Maligned or malignant? *American Psychologist*, *2*, 26–273.

Coleman, J. S. (1961). *The adolescent society: The social life of the teenager and its impact on education*. Glencoe, IL: Free Press.

Côté, S. M., Boivin, M., Nagin, D. S., Japel, C., Xu, Q., Zoccolillo, M., et al. (2007). The role of maternal education and non-maternal care services in the prevention of children's physical aggression. *Archives of General Psychiatry*, *64*, 1305–1312.

Dmitrieva, J., Steinberg, L., & Belsky, J. (2007). Child-care history, classroom composition, and children's functioning in kindergarten. *Psychological Science*, *18*, 1032–1039.

Estell, D. B., Cairns, R. B., Farmer, T. W., & Cairns, B. D. (2002). Aggression in inner-city early elementary classrooms: Individual and peer-group configurations. *Merrill-Palmer Quarterly*, *48*, 52–76.

Fabes, R. A., Hanish, L. D., & Martin, C. L. (2003). Children at play: The role of peers in understanding the effects of child care. *Child Development, 74,* 1039–1043.

Farmer, T. W., Estell, D. B., Bishop, J. L., O'Neal, K., & Cairns, B. D. (2003). Rejected bullies or popular leaders? The social relations of aggressive subtypes of rural African American early adolescents. *Developmental Psychology, 39,* 992–1004.

Farmer, T. W., Leung, M.-C., Pearl, R., Rodkin, P. C., Cadwallader, T. W., & Van Acker, R. (2002). Deviant or diverse peer groups? The peer affiliations of aggressive elementary students. *Journal of Educational Psychology, 94,* 611–620.

Farmer, T. W., Xie, H., Cairns, B. D., & Hutchins, B. C. (2007). Social synchrony, peer networks, and aggression in school. In P. H. Hawley, T. D. Little, & P. C. Rodkin (Eds.), *Aggression and adaptation: The bright side to bad behavior* (pp. 209–233). Mahwah, NJ: Erlbaum.

Gest, S. D. (2006). Teacher reports of children's friendships and social groups: Agreement with peer reports and implications for studying peer similarity. *Social Development, 15,* 248–259.

Gest, S. D., Sesma, A. Jr, Masten, A. S., & Tellegen, A. (2006). Childhood peer reputation as a predictor of competence and symptoms 10 years later. *Journal of Abnormal Child Psychology, 34,* 509–526.

Gresham, F. M., & Elliott, S. N. (1990). *The social skills rating system.* Circle Pines, MN: American Guidance Service.

Hanish, L. D., & Rodkin, P. C. (2007). Bridging children's social development and social network analysis. In P. C. Rodkin & L. D. Hanish (Eds.), *Social network analysis and children's peer relationships* (pp. 1–8). San Francisco: Jossey-Bass.

Haskins, R. (1985). Public school aggression among children with varying day-care experience. *Child Development, 56,* 689–703.

Hawley, P. H. (2002). Social dominance and prosocial and coercive strategies of resource control in preschoolers. *International Journal of Behavioral Development, 26,* 167–176.

Hawley, P. H. (2007). Social dominance in childhood and adolescence: Why social competence and aggression may go hand in hand. In P. H. Hawley, T. D. Little, & P. C. Rodkin (Eds.), *Aggression and adaptation: The bright side to bad behavior* (pp. 1–29). Mahwah, NJ: Erlbaum.

Hawley, P. T., Johnson, S. E., Mize, J. A., & McNamara, K. A. (2007). Physical attractiveness in preschoolers: Relationships with power, status, aggression and social skills. *Journal of School Psychology, 45,* 499–521.

Hays, W. L. (1988). *Statistics* (4th ed.). Fort Worth, TX: Holt, Rinehart & Winston.

LaFontana, K. M., & Cillessen, A. H. N. (2002). Children's perceptions of popular and unpopular peers: A multimethod assessment. *Developmental Psychology, 38,* 635–647.

Lease, A. M., Musgrove, K. T., & Axelrod, J. L. (2002). Dimensions of social status in preadolescent peer groups: Likability, perceived popularity, and social dominance. *Social Development, 11,* 508–533.

Lintong, M., & van IJzendoorn, M. H. (2009). Dissipating or diffusing aggression after non-maternal childcare? Commentary on a novel hypothesis. *Social Development, 18,* 239–246.

Luthar, S. S., & McMahon, T. J. (1996). Peer reputation among inner-city adolescents: Structure and correlates. *Journal of Research on Adolescence, 6,* 581–603.

Maccoby, E. E., & Lewis, C. C. (2003). Less day care or different day care? *Child Development, 74,* 1069–1075.

Masten, A. (2001). Ordinary magic: Resilience processes in development. *American Psychologist, 56,* 227–238.

Milgram, S. (1976). *Human aggression* [Videorecording]. University Park: Pennsylvania State University Audio-Visual Services.

Moore, T. (1975). Exclusive early mothering and its alternatives: The outcome to adolescence. *Scandinavian Journal of Psychology, 16,* 255–272.

Murphy, G., Murphy, L. B., & Newcomb, T. M. (1937). *Experimental social psychology.* New York: Harper & Brothers.

National Institute of Child Health and Human Development Early Child Care Research Network. (1996). Characteristics of infant child care: Factors contributing to positive caregiving. *Early Childhood Research Quarterly, 11,* 269–306.

National Institute of Child Health and Human Development Early Child Care Research Network. (1998). Early child care and self-control, compliance and problem behavior at 24 and 36 months. *Child Development, 69,* 1145–1170.

National Institute of Child Health and Human Development Early Child Care Research Network. (2001). Child care and children's peer interaction at 24 and 36 months: The NICHD Study of Early Child Care. *Child Development, 72,* 1478–1500.

National Institute of Child Health and Human Development Early Child Care Research Network. (2002). Early child care and children's development prior to school entry: Results from the NICHD Study of Early Child Care and Youth Development. *American Educational Research Journal, 39,* 133–164.

National Institute of Child Health and Human Development Early Child Care Research Network. (2003a). Does amount of time spent in child care predict socioemotional adjustment during the transition to kindergarten? *Child Development, 74,* 976–1005.

National Institute of Child Health and Human Development Early Child Care Research Network. (2003b). Families matter—Even for kids in child care. *Developmental and Behavioral Prediatrics, 24,* 58–62.

National Institute of Child Health and Human Development Early Child Care Research Network. (2004a). Trajectories of physical aggression from toddlerhood to middle childhood: Predictors, correlates, and outcomes. *Monographs of the Society for Research in Child Development, 69*(4, Serial No. 278).

National Institute of Child Health and Human Development Early Child Care Research Network. (2004b). *COS qualitative scales and reliabilities observational ratings of the fifth grade classrooms*. Research Triangle Park, NC: RTI. Retrieved June 3, 2008, from https://secc.rti.org/project/display.cfm?t=c&i=ccdr571

National Institute of Child Health and Human Development Early Child Care Research Network. (2005). Early child care and children's development in the primary grades: Follow-up results from the NICHD Study of Early Child Care. *American Educational Research Journal, 42*, 537–570.

National Institute of Child Health and Human Development Early Child Care Research Network. (2008). Social competence with peers in third grade: Associations with earlier peer experiences in childcare. *Social Development, 17*, 419–453.

Olweus, D. (2010). Understanding and researching bullying: Some critical issues. In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *Handbook of bullying in schools: An international perspective* (pp. 9–33). New York: Routledge.

Parkhurst, J. T., & Hopmeyer, A. (1998). Sociometric popularity and peer-perceived popularity: Two distinct dimensions of peer status. *Journal of Early Adolescence, 18*, 125–144.

Pearl, R., Leung, M.-C., Van Acker, R., Farmer, T. W., & Rodkin, P. C. (2007). Fourth- and fifth-grade teachers' awareness of their classrooms' social networks. *Elementary School Journal, 108*, 25–39.

Pellegrini, A. D. (2007). Is aggression adaptive? Yes: Some kinds are and in some ways. In P. H. Hawley, T. D. Little, & P. C. Rodkin (Eds.), *Aggression and adaptation: The bright side to bad behavior* (pp. 85–105). Mahwah, NJ: Erlbaum.

Pellegrini, A. D., Roseth, C. J., Milner, S., Bohn, C. M., Van Ryzin, M., Vance, N., et al. (2007). Social dominance in preschool classrooms. *Journal of Comparative Psychology, 121*, 54–64.

Pepler, D. J., Craig, W. M., & Roberts, W. L. (1998). Observations of aggressive and nonaggressive children on the school playground. *Merrill-Palmer Quarterly, 44*, 55–76.

Prinstein, M. J., & Cillessen, A. H. N. (2003). Forms and functions of adolescent peer aggression associated with high levels of peer status. *Merrill-Palmer Quarterly, 49*, 310–342.

Rodkin, P. C., Farmer, T. W., Pearl, R., & Van Acker, R. (2000). Heterogeneity of popular boys: Antisocial and prosocial configurations. *Developmental Psychology, 36*, 14–24.

Rodkin, P. C., Farmer, T. W., Pearl, R., & Van Acker, R. (2006). They're cool: Social status and peer group supports for aggressive boys and girls. *Social Development, 15*, 175–204.

Rodkin, P. C., & Wilson, T. (2007). Aggression and adaptation: Psychological record, educational promise. In P. H. Hawley, T. D. Little, & P. C. Rodkin (Eds.), *Aggression and adaptation: The bright side to bad behavior* (pp. 235–267). Mahwah, NJ: Erlbaum.

Roisman, G. I., Susman, E., Barnett-Walker, K., Booth-LaForce, C., Owen, M. T., Belsky, J., et al. (2009). Early family and child-care antecedents of awakening cortisol levels in adolescence. *Child Development, 80*, 907–920.

Rose, A. J., Swenson, L. P., & Waller, E. M. (2004). Overt and relational aggression and perceived popularity: Developmental differences in concurrent and prospective relations. *Developmental Psychology, 40*, 378–387.

Schwartz, D., Gorman, A. H., Nakamoto, J., & McKay, T. (2006). Popularity, social acceptance, and aggression in adolescent peer groups: Links with academic performance and school attendance. *Developmental Psychology, 42*, 1116–1127.

Sippola, L. K., Paget, J., & Buchanan, C. M. (2007). Praising Cordelia: Social aggression and social dominance among adolescent girls. In P. H. Hawley, T. D. Little, & P. C. Rodkin (Eds.), *Aggression and adaptation: The bright side to bad behavior* (pp. 157–183). Mahwah, NJ: Erlbaum.

Stormshak, E. A., Bierman, K. L., Bruschi, C., Dodge, K. A., Coie, J. D., & Conduct Problems Prevention Research Group. (1999). The relation between behavior problems and peer preference in different classroom contexts. *Child Development, 70*, 169–182.

Strayer, F. F. (1980). Social ecology of the preschool peer group. In W. A. Collins (Ed.), *Minnesota symposium on child development* (Vol. 13, pp. 165–196). Hillsdale, NJ: Erlbaum.

Sutton, J., Smith, P. K., & Swettenham, J. (1999). Bullying and "theory of mind": A critique of the "social skills deficit" view of anti-social behaviour. *Social Development, 8*, 117–127.

Thompson, R. A. (2009). Making the most of small effects. *Social Development, 18*, 247–251.

Vaillancourt, T., & Hymel, S. (2006). Aggression and social status: The moderating roles of sex and peer-valued characteristics. *Aggressive Behavior, 32*, 396–408.

Vaughn, B. E., Vollenweider, M., Bost, K. K., Azria-Evans, M. R., & Snider, J. B. (2003). Negative interactions and social competence for preschool children in two samples: Reconsidering the interpretation of aggressive behavior for young children. *Merrill-Palmer Quarterly, 49*, 245–278.

Woodcock, R. W. (1990). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment, 8*, 231–258.

Wright, J. C., Giammarino, M., & Parad, H. W. (1986). Social status in small groups: Individual-group similarity and the social "misfit." *Journal of Personality and Social Psychology, 50*, 523–536.

Xie, H., Li, Y., Boucher, S. M., Hutchins, B. C., & Cairns, B. D. (2006). What makes a girl (or a boy) popular (or unpopular)? African American children's perceptions and developmental differences. *Developmental Psychology, 42*, 599–612.

Zeller, M., Vannatta, K., Schafer, J., & Noll, R. B. (2003). Behavioral reputation: A cross-age perspective. *Developmental Psychology, 39*, 129–139.

Child Development, May/June 2010, Volume 81, Number 3, Pages 851–869

The Rise in Cortisol in Family Day Care: Associations With Aspects of Care Quality, Child Behavior, and Child Sex

Megan R. Gunnar, Erin Kryzer,
and Mark J. Van Ryzin
University of Minnesota

Deborah A. Phillips
Georgetown University

This study examined the increase in salivary cortisol from midmorning to midafternoon in 151 children (3.0–4.5 years) in full-time home-based day care. Compared to cortisol levels at home, increases were noted in the majority of children (63%) at day care, with 40% classified as a stress response. Observations at day care revealed that intrusive, overcontrolling care was associated with the cortisol rise. For girls, the cortisol rise was associated with anxious, vigilant behavior, while for boys the rise was associated with angry, aggressive behavior. Child behavior did not mediate or moderate relations between care quality and the cortisol rise, except for evidence that boys scoring low on angry, aggressive behavior were more sensitive to variations in warm, supportive care than boys scoring high on this behavior.

Years of study of the impact of child care on children's development has led to the general conclusion that whereas child care can confer benefits (e.g., Ahnert & Lamb, 2003; Love et al., 2003; Phillips, McCartney, & Sussman, 2006), long hours in care may have some negative impacts that may be more pronounced for experiences in child-care settings of poorer quality (e.g., Belsky et al., 2007; Gunnar & Donzella, 2002; National Institute of Child Health and Human Development Early Child Care Research Network [NICHD ECCRN], 1998, 2001, 2003). Specifically, higher quality care has been associated with more positive peer interactions and lower levels of externalizing behavior (Howes et al., 1996; NICHD ECCRN, 1998, 2001, 2001, 2003; Votruba-Drzal, Coley, & Chase-Lansdale, 2004) and long hours in care are associated with the emergence of behavior problems in childhood (Belsky et al., 2007; NICHD ECCRN, 2001, 2003). Although the association between quantity of care and externalizing behavior is generally maintained in models with controls for family background and child-care quality, a recent re-analysis of data from the NICHD Study of Child Care and Youth Development has documented that this association is moderated by the quality of care children experienced prior to school entry (NICHD ECCRN, 2010). Hours of child care over periods when the child was in care was a stronger predictor of externalizing behavior at 24 and 54 months in lower quality compared to higher quality care.

In attempts to understand the mechanisms through which child care may operate to increase risks of negative developmental outcomes, researchers have begun to explore relations between child care and activity of the hypothalamic–pituitary–adrenocortical (HPA) system. The HPA system produces cortisol, a steroid hormone that plays critical roles in adaptation to stressors (for discussion of the HPA axis and the neurophysiology of stress, see Gunnar & Quevedo, 2007). Chronic or frequent activations of this neuroendocrine system early in life have been shown in animal models to increase fearfulness and impair behavior and physiological regulatory competence (Gunnar & Quevedo, 2007).

Interest in child care as a stressor capable of activating the HPA axis can be traced to a serendipitous finding published a decade ago (Tout, de Haan,

---

This research was supported by the National Institute of Mental Health through a grant (MH62601) and a Senior Scientist Award (MH066208) to Megan R. Gunnar and through a grant provided by the National Institute of Child Health and Human Development (MH165802) to Deborah A. Phillips. We would like to thank Deborah Vandell and Carollee Howes for their assistance in modifying and developing the M-ORCE. Finally, this study would not have been possible without the generous cooperation of the participating family day-care providers and families they serve.

Correspondence concerning this article should be addressed to Megan R. Gunnar, Institute of Child Development, University of Minnesota, Minneapolis, MN 55455. Electronic mail may be sent to gunnar@umn.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8103-0014

Campbell, & Gunnar, 1998). Tout et al. (1998) were studying associations between day-to-day variability in cortisol levels and children's socioemotional behavior. When they examined their morning and afternoon cortisol data, they found that most of the children showed a rise in cortisol from morning to afternoon and they did so on most of the days they were in child care. This was remarkable as the typical basal pattern of cortisol production between midmorning and midafternoon tends to be level or declining for children of this age (Watamura, Donzella, Kertes, & Gunnar, 2004).

Since publication of the Tout et al. (1998) study, there have been a number of studies examining cortisol activity at child care. Two recent meta-analyses have summarized those findings (Geoffroy, Côté, Parent, & Séguin, 2006; Vermeer & van IJzendoorn, 2006). Both papers concluded that the rise in cortisol from morning to afternoon at child care has been convincingly documented and that this rise is not seen in the same children when they are at home on non-child-care days. Both meta-analyses also concluded that age is a relevant factor, with larger increases being observed among the younger children (2- and 3-year-olds) compared to the older children (4- and 5-year-olds). However, the two meta-analyses differed with regard to two critical issues: (a) whether child-care quality influences the magnitude of the cortisol stress response over the day at child care and (b) whether elevations in cortisol at child care are associated with children's temperament or behavior in the care setting.

Regarding quality of care, the first meta-analysis (Geoffroy et al., 2006) concluded that it has been shown conclusively that cortisol levels are higher and rise more over the day in poorer quality child care, whereas the second meta-analysis (Vermeer & van IJzendoorn, 2006) noted that nearly all of the child-care studies of cortisol activity have been conducted in relatively high-quality care settings and thus child-care quality cannot be the primary factor determining increases in cortisol at child care. The Vermeer and van IJzendoorn (2006) analysis was restricted to studies of children in child care, whereas the Geoffroy et al. (2006) analysis included a broader range of social settings. Thus, a conservative conclusion is that the relation between care quality and the cortisol increase over the child-care day has not been fully explored.

Indeed, only a few studies that have directly examined increases in cortisol in relation to measures of child-care quality. Legendre (2003) examined structural characteristics of eight child-care centers and found that both the presence of large (> 15) numbers of children and more than four adults, in addition to a wide age spread, were associated with greater cortisol increases. Sims, Guilfoyle, and Parry (2006) examined child-care settings and grouped them as high quality, satisfactory, and unsatisfactory. Among 3- to 5-year-old children, significant *decreases* in cortisol over the day were noted for high-quality centers, whereas in unsatisfactory centers, significant *increases* were noted. To our knowledge, there has been only one study examining whether care quality is associated with cortisol activity in family day-care settings. Dettling, Parker, Lane, Sebanc, and Gunnar (2000) examined a small number of family day-care settings targeting one child in each setting. They noted that poorer quality of care using items from the Observational Rating of Care Environments (NICHD Early Daycare Research Network, 2000) was associated with larger increases in cortisol, but found no evidence that group size in these day-care homes was a significant factor. Thus, the evidence that quality of child care affects the rise in cortisol at child care is not conclusive, especially with regard to family day care. Furthermore, the association between the facets of quality, including structural characteristics (e.g., group size, age range, ratio of caregivers to child) and child–provider interactions (i.e., process measures of quality) needs further exploration. The first two goals of the study were to (a) determine whether cortisol increases over the day in family day-care settings and (b) determining whether increases over the child-care day in these settings were associated with structural or process measures of child-care quality.

The third goal was to examine relations between child behavior and the rise in cortisol over the child-care day. As noted, the two meta-analyses also differed in their conclusions about child behavior. The Geoffroy et al. (2006) analysis concluded that children with difficult temperaments exhibit higher cortisol levels at child care, whereas the Vermeer and van IJzendoorn (2006) analysis concluded that the association between children's emotional dispositions and the rise in cortisol has yet to be shown conclusively. Again the difference in conclusions lies partly in the fact that the two meta-analyses included different studies, with the Geoffroy et al. analysis including studies of children in settings other than child care. Because the question pertains to child behaviors associated with the rise in cortisol over the child-care day, focusing specifically on full-day child-care settings is essential to answering the question.

Based on the human and animal literature, there are two aspects of child behavior that might impact HPA axis stress responses in child-care settings: anxious, vigilant behavior and angry, aggressive behavior. Regarding anxious vigilance, neuroscience evidence on the neural pathways activating stress hormone reactions documents the importance of input from the amygdala and distributed systems involved in fear and anxious vigilance (for a review, see Rosen & Schulkin, 1998). Thus, it has long been suspected that children who are more anxious and vigilant would be more vulnerable to producing activation of the HPA axis (e.g., Kagan, Reznick, & Snidman, 1987). The neuroscience evidence regarding angry and aggressive behavior is mixed. Whereas there is some evidence that children extreme in these behaviors (e.g., oppositional defiant disorder) have lower than normal basal and response levels of cortisol (for a review, see van Goozen, Fairchild, Snoek, & Harold, 2007), there is also evidence that among preschool-aged children aggressive and under controlled behavior in peer group settings is associated with higher cortisol levels (for a meta-analysis, see Alink et al., 2008). This latter finding may partly reflect cortisol increases in response to peer rejection, which occurs more among children who are aggressive and undercontrolled (Gunnar, Sebanc, Tout, Donzella, & van Dulmen, 2003). Thus, the third goal of the study was to examine how patterns of anxious, vigilance and angry, aggressive behavior at child care might relate to cortisol stress reactions over the day in family day-care settings.

In examining these aspects of child behavior, we were aware that boys and girls might exhibit these anxious, vigilant and angry, aggressive behaviors with different frequencies. In particular, boys have often been noted to be more physically aggressive with peers (Crick et al., 2006; Fabes, Martin, & Hanish, 2003), although girls may exhibit relational aggression somewhat more than boys (Card, Stucky, Sawalani, & Little, 2008; Crick et al., 2006). To the extent that these behaviors reflect challenges the children were having in these settings, then their different prevalence might also be related to gender differences in patterns of association with cortisol increases over the child-care day. Indeed, studies of cortisol–behavior associations for children in group settings do report sex differences in patterns of findings, although the direction of effects is not consistent. Several of the studies report associations for one sex but not the other, and sometimes for different measures for each sex within the same study (Dettling, Gunnar, & Donzella, 1999; Tout et al., 1998). While these sex differences might be spurious, they may also reflect differences in how emotional dispositions influence children's experiences in peer group settings. For example, we have noted that preschool girls who are popular with other girls tend to score higher on measures of negative emotionality and lower on measures of assertiveness, whereas boys who are popular with other boys tend to be more exuberant and higher in assertiveness (Sebanc, Pierce, Cheatham, & Gunnar, 2003). These findings may also reflect differences in how care providers respond to boys versus girls with different temperaments and behavioral styles (Phillips et al., 2010). Thus, the fourth goal of the study was to examine whether gender moderated the relations between child behavior and the child-care cortisol stress response.

Finally, we were also interested in how these patterns of child behavior might operate in relation to child-care quality measures in statistical models predicting the rise in cortisol over the child-care day. Both mediating and moderating relations were explored. Regarding mediating relations, the theoretical model for these relations assumed that in order for care quality to impact activity of the HPA system it must have an emotional-behavioral impact on the child. Furthermore, the aspects of behavior impacted must reflect or connect with neural systems capable of activating the HPA stress response. This model is fairly straightforward for anxious vigilance: Children displaying more anxious, vigilant behavior were expected to be experiencing more activity in cortico-limbic pathways that support fear and vigilance and which, via limbic-hypothalamic connections, converge on the HPA axis to produce stress elevations in cortisol (e.g., Rosen & Schulkin, 1998). Children in poorer quality care settings would be expected to experience more frequent periods of fear and anxiety (see also Kryzer, Kovan, Phillips, Domagall, & Gunnar, 2007) and thus should produce greater rises in cortisol at child care. The model of angry, aggressive behavior, however, is less straight-forward. Here the assumption is that these behaviors, which involve negative social interactions, result in heightened experiences of social threat which also activate amygdalar outputs to the HPA axis and thus produce stress elevations in cortisol at child care (see Dickerson & Kemeny, 2004).

Models predicting moderation assume that children's behavior at child care at least partially reflects traits that may make them more or less vulnerable to stress as child-care quality decreases. For example, Boyce and Ellis (2005) have argued that anxious, vigilant children are more sensitive to context, not only experiencing more stress and negative outcomes when conditions are poor but less stress and better outcomes than other children when rearing conditions are especially supportive. Similarly, Belsky, Hsieh, and Crnic (1998) proposed that infant negative emotionality, which includes proneness to anger, increases children's susceptibility to externalizing problems as a function of differential rearing experiences. Extrapolating from these theoretical models, both anxious, vigilant and angry, aggressive dispositions may moderate associations between care quality and cortisol stress responses.

To summarize, the purpose of this study was to develop a better understanding of the factors associated with rise in cortisol over the child-care day. There were five goals: (a) determine whether cortisol increases of the day in family day-care settings; (b) examine the extent to which structural and process measures of child-care quality are associated with the magnitude of cortisol increase over the child-care day; (c) explore whether anxious, vigilant or angry, aggressive behavior is related to the child-care cortisol stress response; (d) examine whether gender moderates child behavior–cortisol rise associations; and (e) examine whether child behavior mediates or moderates associations between process measures of child-care quality and the child-care cortisol stress response.

In conducting this study, we focused on children in family day-care homes for several reasons. First, relative to center-based care settings, there are less data on whether and to what extent cortisol increases over the day in family-based day care. Second, we were able to collect a large number of day-care homes for analysis, many more than had we tried to sample across different child-care centers, thus broadening our ability to examine relations between care qualities and HPA axis activity. Third, family-based day-care settings range widely in the number of children in care, with some including only one or two children, whereas others operate more like small day-care centers with 10 or more children and several adults, giving us the ability to evaluate the effects of group size. Finally, although we only examined licensed day-care homes, the education levels, day-care experience of the providers, and observational measures of care quality were expected to vary considerably, again allowing a broad examination of quality-related factors that might help explain variation in cortisol increases among the children.

Method

Participants

The children ($N = 151$, 54.3% girls) were recruited from family-based day-care settings in a major metropolitan area. Selection criteria were: child age 3–4.5 years ($M = 3.81$, $SD = 0.23$), full-time care (≥ 20 hr per week; $M = 40.61$ hr, $SD = 8.26$ hr), and at least 2 months attending the present *day-care* home ($M = 27.52$ months, $SD = 13.40$ months, range = 2–48 months). This last criterion was used to avoid assessing children who might still be adapting to the day-care arrangement. Of the children, 85.4% were White, 12% African American, 1% Asian, and 1% Hispanic. Demographic data were collected on the families: Ninety-six percent provided the requested information. Income before taxes was coded in $25,000 increments with mean income in the $51,000–$76,000 range. Parent education was coded as: (1) less than high school, (2) high school or general education diploma (GED), (3) some college, (4) associate's degree, (5) bachelor's degree, (6) master's degree, and (7) professional school or doctorate degree. Median level parental education was an associate's degree, with no parent reporting less than a high school or GED education. Only 4% of the children lived in single parent households. As such, we consider the sample to be low risk.

Recruitment Procedures

Children were recruited through licensed family-based child-care providers who had responded to solicitations for participation mailed directly to them or placed in the licensing agencies' newsletter. Providers were contacted to determine if they had a child in their care that met the selection criteria. Eligible providers who agreed to participate approached the parent(s) of the child to request permission to release their contact information. Parents were then contacted to request consent for their child to participate; of the parents contacted, 59% agreed to participate. Eighty-one participants were from child-care sites that contributed only a single child, whereas 70 were from 30 sites that contributed two or three participants each. Recruitment took place over multiple years, implying that children from the same site were not necessarily present at the same time in the child-care setting. Of the 30 sites with multiple participants, 36 of the 70 participants were present together at the same time. Sites contributing more than one child to the analyses tended to be the larger child-care sites, $t = 1.71$, $df = 149$, $p < .10$.

Day-Care Settings

Day-care providers completed a brief questionnaire about their setting and experience. All 120 day-care sites were licensed. The number of children present ranged from 2 to 12 ($M = 5.25$, $SD = 2.25$) and the number of adults at each site ranged from 1 to 3 ($Mdn = 1$, $M = 1.14$, $SD = 0.44$). The ratio of children to each adult who was physically present and within sight of the children averaged at the time of our observations was 1:5.38 ($SD = 2.25$, range = 1:1–1:25). Ninety-six percent of the providers classified themselves as non-Hispanic White. Providers averaged 13.84 ($SD = 7.69$) years of work as a day-care provider, and 30.7% were providing care to a child of their own who was under the age of 5 years. We note that being the child of the care provider was an *exclusion* criterion. Using the same education scale described above, the median education of the providers was some college, including 2-year degree from a technical or community college. We also examined formal child development or early childhood education course work in high school or college, 11.30% reported no such formal coursework, 52% reported some formal training but not college level, whereas 37% reported college level training.

Procedures and Measures

Care provider and child behavior measures. The children were observed on two mornings between 8:30 and 9:30 a.m. on days the provider deemed *typical* (i.e., no field trip, birthday parties, etc.). The observation inventory developed for this research project was a modification of the Observational Rating Scales of Caregiving Environment (ORCE; see NICHD ECCRN, 2000; for a complete description of the modified or M-ORCE, see Kryzer et al., 2007). Like the ORCE, the M-ORCE involves two 44-min observations (one per day), divided into four 10-min observation periods. In each 10-min period, observers alternate between 30 s observe and record frames for coding frequency counts of behavior. At the end of each 10-min period, the observer makes brief notes and tentative qualitative ratings for 2 min. This process was repeated for three 10-min periods. In the final 10-min period, the observer makes observations exclusively for the qualitative ratings. All qualitative ratings were 4-point scales. Coding of the M-ORCE was performed by six coders, one of whom was the reliability coder who trained the others. The reliability coder scored 41% of the children in the study, while the other coders contributed between 3% and 19% of the data each. Coders were trained through discussion of the codes and practice on videotapes. They then accompanied the reliability coder to practice coding sites and scored with her until they achieved Cohen's kappas of at least .80 on all scales. During data collection, coder agreement was reassessed approximately every fifth session for each coder by having the reliability coder jointly code one of the target subjects with each coder. Approximately every 6 weeks, the coders reviewed a master coding videotape and discussed discrepancies to reduce problems of coding drift.

For the present report, nine qualitative variables of care quality were examined as dimensions of the day-care environment expected to influence the stressfulness of the context. These variables were: Sensitivity, Intrusiveness, Detachment-Disengagement, Positive Regard for the Child, Negative Regard for the Child, Positive Community Building, Negative Community Building, Environmental Rating of Chaos, and Environmental Rating of Overcontrol. Sensitivity, Positive and Negative Regard, Detachment, and Intrusiveness were rated based on the provider's behavior toward the target child. Overcontrol was a rating of the care provider's structuring of the care environment, whereas Chaos reflected the general degree of unpredictability of the setting. These measures were identical on the ORCE and M-ORCE. Positive Community Building was a measure of the care provider's use of strategies to help the children understand and relate positively to one another (e.g., "Johnny is feeling sad today because his mother is in the hospital, so that is why he wants to play alone right now"). Negative Community Building was a measure of the care provider's use of strategies that created winners and losers and children pitted against one another as a discipline strategy (e.g., "Sam isn't doing what I asked, is he? We can't start the story until Sam is quiet"). These measures were developed based on work by Lewis, Watson, and Schaps (2003). As would be expected, Positive and Negative Community Building were very highly correlated ($r > .80$) and thus were combined into one scale by subtracting Negative from Positive Community Building scores.

To reduce the number of scale examined, the eight qualitative measures noted above were entered into a principal axis factor analysis. The results indicated two factors with eigenvalues > 1 that accounted together for 68% of the variance. The first factor, with an eigenvalue of 3.98, explained 49.7% of the variance and consisted of Sensitivity, Detachment (negative loading), Positive Regard, Chaos (negative loading), and the summary Community Building variable. The resulting measure was labeled Warm-Supportive Care. The second factor with an eigenvalue of 1.5 accounted for 19% of the variance and consisted of Intrusiveness, Negative Regard, and Overcontrol. We labeled the resulting measure Intrusive-Overcontrolling Care. After varimax rotation, there were no variables that cross-loaded > 0.40.

Six measures of child functioning were examined. Three were qualitative measures: Vigilant-Anxious Mood, Angry-Irritable Mood, and Belonging. Three were based on the frequency counts: Solitary Play, Positive Integration, and Negative Peer Interactions. All except Solitary Play were measures developed for the M-ORCE. Vigilant Mood (1 = *no evidence of Vigilant-Anxious behavior*, 4 = *high level of Vigilant-Anxious behavior*) described the extent to which the child engaged in such behaviors as flinching when others came near, moving quickly away from others, and continuously dividing attention between what they were doing and tracking events in the environment. Angry-Irritable Mood (1 = *no evidence of Angry-Irritable behavior*, 4 = *high level of Angry-Irritable behavior*) described the extent to which the child expressed anger and irritability in face, voice, or gesture. While such behavior might involve interactions with others, it could also reflect behavior directed toward objects (e.g., throwing down a toy). Belonging was a qualitative rating reflecting the extent to which the child acted relaxed, confident, and assured (4 = *acts highly relaxed, confident, and assured*) as opposed to anxious, tentative, and cautious (1 = *anxious, tentative, and cautious*), that is, the extent to which the child seems to "fit in" in the setting as opposed to appearing like an outsider who is not integrated into the setting. Children did not have to be involved in play with other children in order to score high on this measure, as long as their behavior was relaxed and they appeared confidently engaged in what they were doing. Solitary play was a frequency count of the number of 10-s periods when the child was playing alone computed as the percentage of intervals when the child was not engaged in physical care activities (i.e., was available for social interaction). Positive Integration assessed not only the frequency with which the child was positively engaged with others in the setting, but also the extent to which the child's role in the interaction was central to the group. A score of 4 indicated positive engagement where the child was a key player in the interactions, directing the interactions or being looked to by others while a lower score indicated positive engagement but in a role where the child was a subordinate or more peripheral actor in the activities. Thus this measure reflected not only the amount but the centrality of the child in positive social interactions. Negative Peer Interactions was a frequency count of the number of intervals in which the child was either the recipient or agent of aggressive behaviors towards other children. Aggression was scored as verbal, physical, and relational aggressive acts. Verbal acts included verbal threats of physical harm, physical acts included pushing and shoving as well as hitting and forcibly attempting to take things away from other children, while relational aggression involved threats or actions directed at harming relationships with peers (e.g., "If you don't give me that, I won't be your friend"; see Crick, Casas, & Mosher, 1997).

These child measures were also submitted to a principal component analysis to permit data reduction. The initial results with varimax rotation indicated that Solitary Play loaded greater than .40 on multiple factors and thus was removed. The resulting analysis yielded two factors with eigenvalues > 1 that accounted for 73% of the variance. The first factor had an eigenvalue of 2.52 and accounted for 50% of the variance. It consisted of Belonging, Positive Integration, and Vigilant-Anxious Mood (negative loading). The resulting measure was reverse scored to yield a measure of Anxious, Vigilant Behavior. The second factor with an eigenvalue of 1.13 accounted for 23% of the variance and consisted of Angry-Irritable Mood and Negative Peer Interactions. The resulting measure was labeled Angry-Aggressive. After varimax rotation, none of the variables cross-loaded > .40.

To examine the stability or trait-like aspect of these measures over time, we examined data on a subset of the children ($N = 108$) who were seen in these same child-care settings 6 months later. The test–retest correlations computed on these participants were: Warm, Supportive Care, $r = .49$, $p < .001$; Intrusive, Overcontrolling Care, $r = .37$, $p < .001$; Anxious, Vigilant, $r = .29$, $p < .01$; and Angry-Aggressive Behavior, $r = .30$, $p < .01$.

Salivary cortisol. Saliva was obtained for cortisol determination by having the children dip a 1.5-in. cotton dental roll into approximately .025 g of cherry flavored Kool-Aid™ mix and mouth the cotton to obtain the sweet taste. This small amount of Kool-Aid™ has not been found to significantly affect the cortisol assay (Talge, Donzella, Kryzer, Gierens, & Gunnar, 2005). Once the cotton roll was saturated, it was placed in a needleless syringe and the saliva was expressed into a 1.50-ml Eppendorf Safe-Lock microtube and sealed. At day care, samples were collected by the day-care provider on 2 days between 10:00 and 11:00 a.m. and 3:00 and 4:00 p.m. These times were chosen based on previous work showing that when four samples of cortisol are obtained (midmorning, noon, immediately postnap, late afternoon), the rise in cortisol over the child-care day can be reliably assessed at these two time points (Watamura, Sebanc, & Gunnar, 2002). Providers were asked to avoid sampling immediately before a meal, to not give the child caffeinated drinks or dairy products within an hour of sampling, and to wait until 30 min after the child got up from a nap to sample to avoid the decrease in cortisol typically seen over nap time (Watamura et al., 2002). To determine the actual time of sampling, the cotton dental rolls were supplied in a bottle with a MEMS V Track Cap™ (Aardex, Zug, Switzerland), which automatically recorded the time when the container was opened. Use of such devices allows verification of compliance with sampling protocols and also increases compliance (Kudielka, Broderick, & Kirschbaum, 2003). The care provider also completed a brief diary on each day of sampling recording timing of sample, nap times, and meal times.

The data from the MEMs cap and diary were examined to determine if the care provider had complied with the instructions. Twenty children were removed prior to assaying saliva for cortisol because the care provider deviated by more than 1.5 hr in the timing of nearly all of their samples, resulting in the child having no usable cortisol data. The 151 with usable samples are described in this report. Parents were also asked to collect samples using the same protocol on 2 days when the child did not attend day care. These measures provided a home baseline against which to judge the impact of day care on children's cortisol activity. Parents followed the same protocol as the day-care providers, including use of the MEMS V Track Cap™. There were 108 children with a sufficient number of home samples for analysis. Comparisons of child-care cortisol for those with and without home cortisol measures yielded no significant (all $ps > .10$) differences for morning or afternoon levels or for change over the day.

Once collected, day care and home samples were stored in the refrigerator and then mailed to the laboratory. These procedures have been shown not to affect cortisol data (Clements & Parker, 1998). Once in the laboratory, samples were frozen at −20 °C until assay. Assays were conducted in duplicate using a time-resolved fluorescence immunoassay. Intra- and interassay coefficients of variation were at or less than 6.7% and 9.0%, respectively, and duplicates correlated highly, $r = .997$, $p < .001$.

Both home and day-care cortisol measures were averaged within time and context over the 2 days of collection. Within each context, children needed at least one data point at each time of day to obtain a final score. Four measures were computed: Home AM, Home PM, Daycare AM, and Daycare PM. The distributions of these variables were examined and were found to be positively skewed and thus were $\log_{10}$ transformed for analysis. The raw AM values were subtracted from the raw PM values to produce Home Rise and Daycare Rise. These values had roughly normal distributions and thus were not $\log_{10}$ transformed. To examine stability in the rise in cortisol across the days, we also calculated the rise on each of the 2 days of assessment. Finally, to determine whether the levels at child care represented a biologically significant secretory response of the HPA axis for each child, we used the criteria of an increase of at least 0.10 μg/dl and 10% derived from work on previously used in adult studies (Kirschbaum, Pirke, & Hellhammer, 1993; Kirschbaum et al., 1995; Van Cauter & Refetoff, 1985).

Analysis plan and preliminary analyses

In some instances, more than one child was obtained from the same day-care setting, so we accounted for this nesting in all analyses. This was accomplished in SAS using procedures for analyzing nested data (SAS Institute, 2004). In our regression analyses (SAS Proc Surveyreg), the day-care provider was declared to be the "primary sampling unit" or PSU, while in the analysis of variance (ANOVA; SAS Proc Genmod or Proc Mixed) the day-care provider was declared to be the subject. In addition to accounting for the variance associated with children nested within day-care provider, these procedures also adjust the degrees of freedom such that they are based upon the number of day-care providers ($N = 112$) rather than the number of children.

Because some studies of child behavior relations with cortisol activity at child care have yielded evidence of associations for one sex but not the other, in all of the analyses described below we included child sex and examined possible main and interaction effects. The first analyses were conducted to determine whether cortisol rose more over the day at child care than at home. This analysis was computed using a 2 (child sex) × 2 (place: home vs. child care) × 2 (time: morning vs. afternoon) doubly repeated measures ANOVA. We also examined the day-to-day stability of the rise at child care and relied on descriptive statistics to help interpret the percentage of children exhibiting biologically significant increases over the day.

We then attempted to identify background and child-care structural quality factors that contributed to the rise at day care. This analysis included child age and sex, family demographics (e.g., family income), day-care descriptors (e.g., ratio of providers to children), and provider variables (e.g., education, years of experience). Any significant associations were then included as covariates in subsequent analyses of the association of provider behavior measures of quality with the cortisol rise at child care. Next, we examined the two care day-care observational quality measures and the two child behavior measures. Initial analyses examined possible age and sex effects for each variable. Subsequent analyses examined first the association of observed quality measures and then child behaviors with the rise in cortisol over the day-care day.

Mediation analyses require that care quality and child behavior measures were associated with the outcome (i.e., Daycare Rise) and, further, that the care quality and child behavior measures were significantly associated. To prepare for mediation analyses, we next examined whether the care quality measures were associated with the child behaviors. Finally, for the measures meeting the requirements for meditational analysis, we constructed a hierarchical regression, entering first the care quality measure(s) associated with the cortisol rise and, in the next step, the child behavior measure(s) associated with the cortisol rise. The question addressed was whether the association with care quality decreased significantly once the child behavior measures were in the equation.

Moderation analyses were computed for child behaviors exhibiting significant associations with the Daycare Rise. For such measures, hierarchical regressions were computed entering each care quality measure and then, in the interaction step, the interaction between each care quality measure and the child behavior measure. Of interest was whether the interaction term produced a significant increase in the variance explained.

Results

Cortisol Change at Day Care and Home

The 2 × 2 × 2 ANOVA of the home and day-care cortisol measures revealed that neither the effect of child sex nor interactions with child sex was significant. However, the time-of-day by place interaction effect was significant, $F(1, 107) = 5.12$, $p < .05$. Descriptive data are shown in Table 1. Tests for simple effects of time-of-day within place were computed to explicate the interaction. At home, we found no significant change from morning to afternoon, $F(1, 107) = 1.12$, *ns*, while at child care, cortisol levels increased significantly from morning to afternoon, $F(1, 107) = 29.42$, $p < .001$. In addition, cortisol levels were higher when the children were at family-based child care than home both in the morning, $F(1, 107) = 9.65$, $p < .01$, and afternoon, $F(1, 107) = 3.92$, $p < .001$, although the effect size was larger in the afternoon ($\eta^2_{\text{a.m.}} = .08$ vs. $\eta^2_{\text{p.m.}} = .24$ and $\omega^2_{\text{a.m.}} = .07$ vs. $\omega^2_{\text{p.m.}} = .23$).

To examine the day-to-day stability of the rise in cortisol at day care, we examined each day separately. At child care, the rise over the day was stable from Day 1 to Day 2, $\beta = .43$, $p < .001$, but this was not the case for the rise at home, $\beta = .04$, *ns*. We then examined the individual values to provide a more descriptive interpretation of these effects. At day care, 63% of the children showed positive (> 0) change from morning to afternoon cortisol levels; 52% showed this degree of rise on each of the 2 days of testing. For those with home data, we examined whether, relative to home baseline, the child's cortisol levels at child care met criteria for a stress response (for criteria, see Kirschbaum et al., 1993). In the morning, 10% of the children were exhibiting an HPA stress response to day care, while in the afternoon this increased to 40%. Examined across the 2 days of assessment, 20% exhibited an HPA stress level response on both of the days, while 56% exhibited such a rise on at least 1 of the 2 days.

Table 1
Means and Standard Deviations for Cortisol Data in μg/dl

Variable	*N*	*M*	*SD*
Home cortisol (10 a.m.)	113	0.12	0.14
Home cortisol (4 p.m.)	112	0.14	0.15
Daycare cortisol (10 a.m.)	151	0.13	0.11
Daycare cortisol in (4 p.m.)	151	0.19	0.18
Daycare cortisol Rise in μg/dl	151	0.06	0.14

Note. Cortisol variables are presented in a linear scale, not log10.

Rise at Day Care Related to Child Demographics and Structural Aspects of Child-Care Quality

Our next step was to relate individual differences in the rise in cortisol over the day-care day to child and caregiver factors. A hierarchical regression predicting Daycare Rise was computed. In Step 1, we included child (i.e., age, sex) and family (i.e., parental education, family income) variables, but none of these measures significantly predicted the rise (*t*s < 1.7 or > −1.7, *ns*). At Step 2, we included structural characteristics of the day-care setting (i.e., number of months at that day-care home, hours per day, group size, number of care providers, ratio of adults to children, and whether a sibling was at that day-care home, with the last entered as a dummy variable). None of these descriptive day-care measures significantly predicted the rise (*t*s < 1.0 or > −1.0, *ns*). At Step 3, we examined day-care provider information (i.e., years of experience, highest level of education, formal training in child development or early childhood education, and whether the provider had their own child under age 5 in the day care, the last coded as a dummy variable). None of the provider variables significantly predicted the rise (*t*s < 1.3 or > −1.3, *ns*). Thus, there was no evidence that any of these child, family, or day-care characteristics affected the magnitude of the rise in cortisol at day care.

We were particularly struck by the lack of effect of the number of children in care as in some of these day-care homes there were only one or two children in addition to the target child, while in others there were 10 or more children. To be sure that there was not a threshold effect of day-care size, we identified 4 children who were in care with only 1 other child (small group) and compared their rise over the day with children in care with 2–5 (medium group) and 6+ other children (large group). Examination of the medians provided no evidence of a differential rise for these three group sizes (i.e., 0.07, 0.02, 0.04 μg/dl, *Mn SD* = 0.03 μg/dl, respectively).

Sex and Age Differences in Child Behavior Measures and Observed Care Quality Measures

Prior to examining associations between process measures of quality, child behavior and cortisol increases, we examined whether these measures differed by child sex or age. Descriptive data on scales that contributed to observed quality measures and child behaviors measures are shown in Table 2. To examine possible gender differences, ANOVAs were computed on the two process measures of quality and the two child behavior measures. The results indicated that girls received more warm, supportive care than did boys (girls: *M* = 0.70, *n* = 82; boys: *M* = −0.83, *n* = 69; *z* =2.25, *p* < .05), while girls exhibited less anxious, vigilant behavior (girls: *M* = −0.47, *n* = 82; boys: *M* = 0.56, *n* = 69; *z* = 2.65, *p* < .01) and less angry, aggressive behavior (girls: *M* = −0.49, *n* = 82; boys: *M* = 0.47, *n* = 69; *z* = 3.39, *p* < .001) than did boys. There were no significant child sex differences in negative, intrusive care (*z* = 1.48, *ns*). We also examined associations with age for the child behavior measures but none were found. The pattern of gender differences supported our concern with examining whether child sex moderated any of the associations in observed measures with the cortisol rise at day care.

Table 2
Descriptive Data on M-ORCE Measures (N = 151)

	M	*SD*
Care Provider Warm-Supportive Care summary score measures		
Sensitivity	3.05	0.39
Detachment (R)	1.58	0.75
Positive Regard	2.68	0.80
Chaos (R)	1.54	0.74
Community Building	0.43	1.27
Care Provider Intrusive-Overcontrolling Care summary score measures		
Negative Regard	1.17	0.45
Intrusiveness	1.51	0.77
Overcontrol	1.55	0.73
Child Anxious-Vigilant summary score measures		
Vigilant, Anxious	1.42	0.65
Belonging (R)	3.03	0.76
Positive Social Integration (R)	31.49	10.88
Angry-Aggressive summary score measures		
Angry-Irritable	1.44	0.69
Negative Peer Interactions	1.51	2.00

Note. R indicates the measure loaded negatively on summary scale.

Day-Care Cortisol Rise Associated With Process Measures of Care Quality

We first examined whether process measures of child-care quality predicted the rise in cortisol using a hierarchical regression; we first entered child sex, then the two measures of quality, and finally the interactions between each observed measure of quality and child sex. As expected, given the null results for gender differences in the

ANOVA of the cortisol child-care rise, in the first step child sex was not significant ($t = 0.25$, *ns*, $R^2 = .00$). In the second step ($\Delta R^2 = .04$), the regression coefficient for Intrusive, Overcontrolling Care was significant ($\beta = .17$, $t = 2.17$, $p < .05$), while Warm, Supportive Care was not ($t = -0.39$, *ns*). In the final step, no interaction terms were significant ($ts < 1.0$, $\Delta R^2 = .00$). Thus, cortisol rose more over the day in family day-care homes where children received more intrusive, overcontrolling care.

Associations Between Child Behavior and Cortisol Rise

Using hierarchical regression, we entered child sex at Step 1, each of the two child behavior measures on Step 2, and the interaction between sex and each of these measures on Step 3 to predict the cortisol rise. The results, shown in Table 3, indicated child sex moderated the associations between both child behavior measures and the cortisol rise. To visualize these interactions, Daycare Rise was plotted by sex for the bottom and top quartiles of the distributions for each child behavior. As shown in Figure 1a, for girls but not boys, Anxious, Vigilant Behavior was associated with larger rises in cortisol over the day-care day. As shown in Figure 1b, for boys, Angry, Aggressive Behavior was associated with the rise in cortisol.

Associations Between Provider Quality Measures and Child Behavior

To prepare for examinations of mediation and moderation, we next examined associations between observed care quality and child behavior. We computed two hierarchical regression models, one predicting Anxious, Vigilant Behavior and the other predicting Angry, Aggressive Behavior. In each we entered child sex on Step 1, the observed care quality measures on Step 2, and the interaction of child sex and the two observed quality measures on Step 3. The results are shown in Table 4. As expected from the ANOVA analysis, Step 1 (child sex) was significant in each equation. For Anxious, Vigilant Behavior (top of Table 4), no significant interactions with sex were observed. However, the second step was highly significant and showed a significant negative association between Warm, Supportive Care and Anxious, Vigilant Behavior; no significant association with Intrusive, Overcontrolling Care was obtained. For Angry, Aggressive Behavior, the interaction step was significant. With the interactions in the equation, Warm, Supportive Care was associated with less Angry, Aggressive Behavior for both sexes, while Intrusive, Overcontrolling Care showed an interaction with child sex. Plotting the top and bottom quartile on Angry, Aggressive Behavior against Intrusive, Overcontrolling Care for each sex (see Figure 2) revealed that boys but not girls exhibited more Angry, Aggressive Behavior in association with higher Intrusive, Overcontrolling Care.

Table 3
Regression Analysis Predicting the Rise in Cortisol as a Function of Child Sex, Child Behavior, and Their Interactions

	Step 1		Step 2		Step 3	
	B	*SE B*	*B*	*SE B*	*B*	*SE B*
Child Sex	.04	.14	.11	.15	.09	.15
Anxious-Vigilant	—	—	.19*	.09	−.01	.10
Angry-Aggressive	—	—	−.01	.12	.15	.15
Sex × Anxious-Vigilant	—	—	—	—	.39*	.18
Sex × Angry-Aggressive	—	—	—	—	−.41*	.20

Note. R^2 (Step 1) = .00; ΔR^2 (Step 2) = .04; ΔR^2 (Step 3) = .06; Total R^2 = .10. Only the highest order terms in each equation are evaluated for significance.
*$p < .05$.

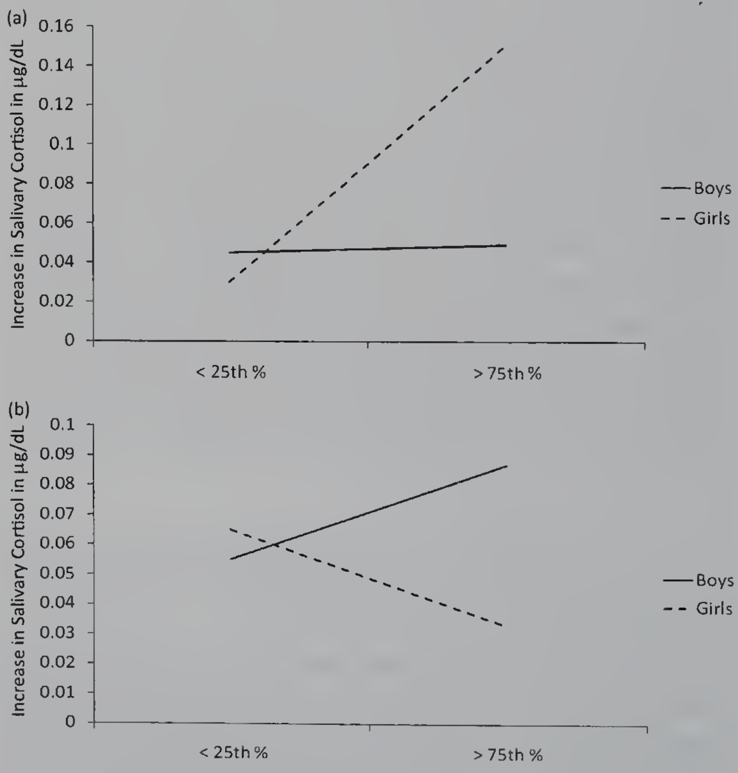


Figure 1. Rise in cortisol over the day-care day as a function of child sex and child behavior. (a) Cortisol rise for the 25th and 75th percentiles of anxious, vigilant behavior. (b) Cortisol rise for the 25th and 75th percentiles of angry, aggressive behavior.

Table 4

Regression Predicting Child Behavior as a Function of Child Sex, Care Provider Behavior, and Their Interactions

	Step 1		Step 2		Step 3	
	B	*SE B*	*B*	*SE B*	*B*	*SE B*
Model 1: Anxious-Vigilant						
Child Sex	−.42**	.16	−.27	.14	−.26	.14
Warm-Supportive	—	—	−.42***	.08	−.35	.12
Intrusive-Overcontrol	—	—	.06	.09	.11	.12
Sex × Warm-Supportive	—	—	—	—	−.14	.17
Sex × Intrusive-Overcontrol	—	—	—	—	−.11	.17
Model 2: Angry-Aggressive						
Child Sex	−.56***	.17	−.43**	.15	−.43	.14
Warm-Supportive	—	—	−.46***	.09	−.44	.12
Intrusive-Overcontrol	—	—	−.08	.09	.17	.12
Sex × Warm-Supportive	—	—	—	—	−.04	.17
Sex × Intrusive-Overcontrol	—	—	—	—	−.52**	.16

Note. Model 1: R^2 (Step 1) = .04; ΔR^2 (Step 2) = .20; ΔR^2 (Step 3) = .00; Total R^2 = .24. Model 2: R^2 (Step 1) = .08; ΔR^2 (Step 2) = .17; ΔR^2 (Step 3) = .07; Total R^2 = .32. Only the highest order terms are evaluated for significance. When the Sex × Warm-Supportive interaction is removed, both individual factors are significant.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

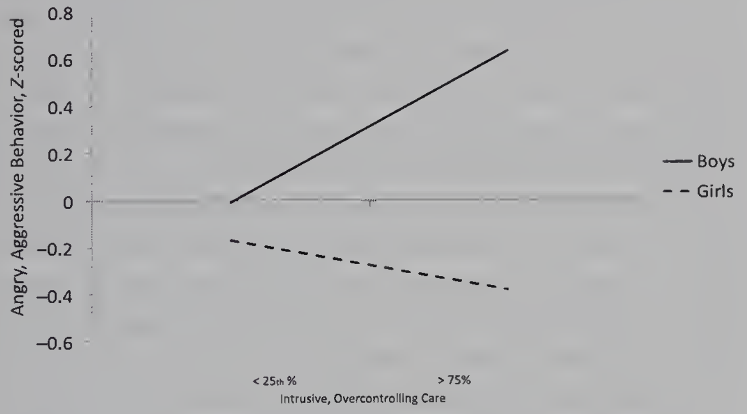


Figure 2. Standardized measure of angry, aggressive behavior as a function of child sex and intrusive, overcontrolling care is plotted using the 25th and 75th percentiles of intrusive, overcontrolling care.

Mediational Analysis

According to Baron and Kenny (1986), we have documented three of the four steps that are required to demonstrate mediation for Intrusive, Overcontrolling Care and Angry, Aggressive Behavior for boys, but not girls. Specifically, we have shown that (a) Intrusive, Overcontrolling Care was related to Daycare Rise; (b) Intrusive, Overcontrolling Care was related to the mediator, Angry, Aggressive Behavior for boys; and (c) Angry, Aggressive Behavior for boys was related to the Daycare Rise. The interaction between Angry, Aggressive Behavior, and sex was related to the outcome (i.e., the rise in cortisol). To examine mediation we examined only the data for boys and computed another hierarchical regression, entering child sex on Step 1, Intrusive, Overcontrolling Care on Step 2, and then Angry, Aggressive Behavior and its interaction with sex on Step 3. Of interest was the decrease in the regression coefficient for Intrusive, Overcontrolling Care when the child behavior scores were in the equation. The results failed to indicate any decrease in the regression coefficient for Intrusive, Overcontrolling Care when Angry, Aggressive Behavior was in the equation; thus, there was no basis for examining whether the decrease was significant (i.e., via a Sobel test).

Moderation Analyses

Because of the gender differences in the pattern of child behavior associations with the cortisol rise, we computed these moderation analyses within sex. For girls, because Anxious, Vigilant but not Angry, Aggressive Behavior was associated with the cortisol rise, we examined whether this behavior moderated associations between observed quality measures and the rise in cortisol. A hierarchical regression was computed entering Anxious, Vigilant Behavior on Step 1, the two observational measures of care quality on Step 2, and the interaction between Anxious, Vigilant Behavior and each of the care quality measures on Step 3. In Step 3 (ΔR^2 = .02), neither the Vigilance × Warm, Supportive Care ($t = -0.26$, *ns*) nor Vigilance × Intrusive, Overcontrolling Care ($t = 1.05$, *ns*) interactions were significant. Thus, there was no evidence that Anxious, Vigilant Behavior moderated the effects of the measures of care quality for girls.

For boys, because Angry, Aggressive Behavior but not Anxious, Vigilant Behavior was associated with the cortisol rise, we examined whether Angry, Aggressive Behavior moderated the associations between observed care quality and cortisol. Again, a hierarchical regression equation was computed, entering Angry, Aggressive Behavior on Step 1, the two measures of care quality on Step 2, and the interaction of Angry, Aggressive Behavior and each care quality measure on Step 3. In Step 3, the interaction between Intrusive, Overcontrolling Care and Angry, Aggressive Behavior was not significant ($t = -0.95$, *ns*, $\Delta R^2 = .00$); however, the interaction between Warm, Supportive Care and Angry, Aggressive Behavior was significant ($t = 2.06$, $p < .05$, $\Delta R^2 = .02$). The interaction is depicted in Figure 3. Boys who were low in aggressive behavior appear to have been sensitive to variations in warm, supportive care, exhibiting larger cortisol increases when they received low levels of supportive care and smaller cortisol increases when they received more supportive care. In contrast, highly aggressive boys appear to exhibit comparable increases in cortisol regardless of the warmth and supportiveness of the provider.

Discussion

The results of the present study confirmed that even in family-based day-care settings, cortisol levels rise over the day. We also found that certain dimensions of care quality and child behavior were associated with the cortisol rise. Finally, although boys and girls did not differ in their cortisol responses to child care, the associations between child behavior, care quality, and the cortisol rise were moderated by child sex. Thus, child sex appears to be a critical factor in understanding the psychosocial processes underlying cortisol stress reactions at child care. Each of these aspects of the results will be discussed below.

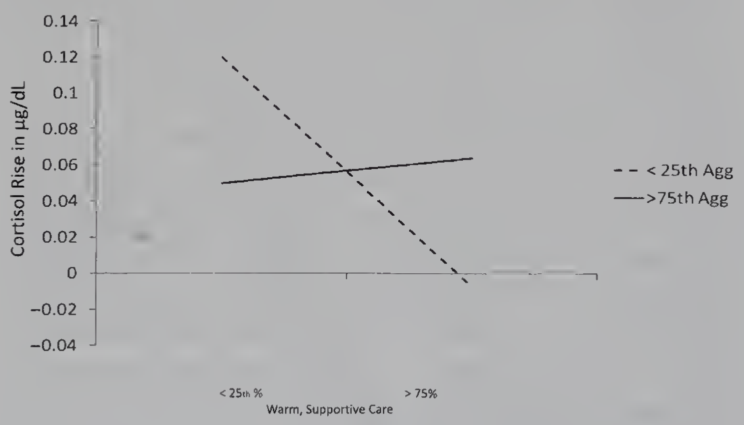


Figure 3. For boys, the interaction between warm, supportive care and angry, aggressive behavior is plotted using the 25th and 75th percentiles of each measure.

The majority of the children showed an increase in cortisol from midmorning to midafternoon at day care, while at-home levels remained relatively flat over the midsection of the day. There was no evidence of a sex difference; boys and girls exhibited similar increases in cortisol over the child-care day. Furthermore, the lack of a change from morning to afternoon in the home data was the typical diurnal basal pattern for children of this age (Watamura et al., 2004). Specifically, the basal decline from midmorning to midafternoon has been shown to emerge at home as circadian sleep patterns become more adult-like (i.e., as the child gives up the afternoon nap). In addition, the rise at day care was relatively stable across the 2 days of assessment, with over half of the children producing a rise from morning levels on each day of assessment. Because even small changes in cortisol could be statistically but not necessarily *biologically* significant, we used criteria previously used in adult studies to identify HPA stress responders. This analysis indicated that only 10% of the children exhibited a stress response in the morning at day care, while 40% could be classified as stress responders in the afternoon.

Taken together these data suggest that a substantial number of children exhibit a stress response of the HPA system over the day at day care; further, as evidenced by the day-to-day stability data, this occurs on a regular basis for a number of children. It is telling that this response is seen more clearly in the afternoon than the morning, indicating a rising pattern of cortisol over the day. This suggests that the response is not a reaction to arriving at child care in the morning, but a response that emerges over the accumulation of the day's experiences. This may explain why in half-day, preschool programs, cortisol levels are not elevated over home levels for classes that meet in the morning or for those meeting in the afternoon (Gunnar, Tout, de Haan, Pierce, & Stansbury, 1997), while these elevations are observed by afternoon in full-day programs. The question, of course, is *why*. What is it about spending a full day at child care that produces a stress response of the HPA system by the afternoon in so many young children?

We found no significant associations between any of the structural measures of care quality and the rise in cortisol at day care. Thus, there was no evidence that provider experience or training predicted this rise, nor was there evidence that the rise was related to the number of children in care or the number of adults in the setting. Our failure to find an association with group size was particularly noteworthy. Not only was no linear association found, but when we examined settings with only one or two children in addition to our target child, increases in cortisol over the day averaged the same as increases in settings with 10 or more children.

Previously we have suggested that the complexity of the demands of managing interactions with large numbers of children might account for cortisol increases over the day among children in center-based care settings (Gunnar & Donzella, 2002; see also Belsky et al., 2007, regarding the possible importance of the peer group in day-care effects). This argument was predicated on the following findings. In comparing the cortisol response over the child-care day for toddlers and preschoolers in center-based care, we did not observe elevations for children cared for in infant rooms (Watamura, Donzella, Alwin, & Gunnar, 2003). In infant rooms, activities are typically organized around the infant's schedule, which means that at any given time, only a few infants are awake and when awake, they are likely to be interacting with adults. In toddler rooms, however, the children eat, sleep and play at the same times and thus there is much more peer interaction in these rooms. Furthermore, with development over the toddler period, time spent interacting with peers increases (Hughes & Dunn, 2007). Second, when we examined changes in cortisol over nap time for preschoolers in center-based care, we noted significant decreases from before to immediately after the nap, even for children who did not appear to sleep over the nap period (Watamura et al., 2002). Because the nap period at child care is one during which children usually do *not* interact with other children, the removal of social interaction demands during this period seemed a reasonable hypothesis for why even those children who did not sleep showed decreases in cortisol over the nap period.

The fact that we observed the same cortisol rise in settings with only two or three children that we found in settings with 10 or more children strongly argues *against* the social complexity hypothesis, at least with regard to day-care settings. It still might be that center-based child-care rooms with many children and adults evoke more of a cortisol response than those with fewer children (see Legendre, 2003), given that none of our settings were that large. However, the fact that the rise in cortisol by afternoon was observed in very small settings suggests that social complexity is not the primary factor accounting for the cortisol stress response. This conclusion also is consistent with the fact that the present findings are largely consistent with previous findings for children of this age studied in center-based care. The average number of children per each preschool-aged care room in day-care centers ranges between 13.0 (at 36 months; NICHD ECCRN, 1999) and 15.4 (at 54 months; NICHD SECCYD, 2008), while the mean number of children in the family day-care homes we studied was around 5. One would have expected, therefore, that if the demands of negotiating complex social environments was the key factor, then many fewer children would have shown a cortisol stress response over the day in family-based day care than has been observed in studies of center-based care. But that was not the case.

We did obtain evidence that our measures of care provider behavior, sometimes termed *process* measure of care quality, were related to the rise in cortisol at day care. Principal factor analysis yielded evidence of two factors: warm and supportive care and intrusive, overcontrolling care. Child-care studies typically identify one dimension of quality that includes both warmth and control dimensions (e.g., NICHD ECCRN, 2000). As in many studies of parenting (e.g., Barber, Stolz, & Olsen, 2005), these dimensions were distinct in our data. It may be that the use of family day-care settings that were more home-like produced patterns of provider behavior that were more like those seen in parenting research.

Our results also showed that these two dimensions were differentially related to the rise in cortisol at day care. Specifically, although warm, supportive care was related to our measures of child behavior, it was not associated with the cortisol rise either for boys or girls (although see discussion of moderation by angry, aggressive behavior for boys, below). In contrast, intrusive, overcontrolling care was associated with larger rises in cortisol over the day-care day. This was true for both boys and girls as gender did not interact with this effect. It is not clear why intrusive, overcontrolling care was stress provoking. It may be that this dimension of care reflected a day-care environment that was structured to provide fewer developmentally appropriate experiences. Anecdotally, settings scoring high in intrusive, overcontrolling care were often ones in which the children were transitioned frequently between activities, were permitted relatively little time in free play, and spent long periods in provider-directed structured activities. In the Sims et al. (2006) study, developmentally appropriate practices related to transitions and scheduling were facets of quality associated with the rise in cortisol for children in child-care centers. It may be that when young children are required to manage a day with many structured activities and transitions, this overtaxes their coping capacities as the day progresses. Notably, however, although intrusive, overcontrolling care accounted for statistically significant variance in the cortisol rise, there was still a good deal of individual variation that was not explained by this process dimension of quality.

Some of this unexplained variance was associated with child behavior patterns at day care. As expected, two dimensions of behavior were identified. The first dimension involved variation in anxiety and vigilance and in social integration. Children scoring high on this dimension were more anxious and vigilant and were less socially integrated in the setting. The second dimension involved variation in angry mood and negative interactions with peers. Both anxious vigilance and angry, aggressive behavior were observed more in boys than girls. Given the association with the cortisol rise, this would suggest that boys might have experienced more stress at child care than girls. However, as noted above, we found no gender difference in cortisol at day care. Rather, we found that gender moderated the associations of behavior and rise in cortisol over the day. For boys, larger cortisol increases were associated with more angry, aggressive behavior, while for girls it was associated with more anxious, vigilant behavior.

Other studies of child care have also noted gender differences in associations between child behavioral dispositions and HPA activity. For example, Dettling et al. (1999) found that parent-reported surgency positively predicted the cortisol rise for boys but not girls, while parent-reported effortful control negatively predicted the rise for girls. Why associations might differ for boys and girls is unclear. By adulthood, there are clear differences in reactivity of the HPA axis to stressors, as well as evidence that the type of stressors that most strongly affect men differ from those that are most provocative for women (for a review, see Dedovic, Wadiwalla, Engert, & Pruessner, 2009). For both men and women, threats to the social self stimulate the HPA axis, but for women the most salient threats to the social self appear to be ones that threaten relationships, while for men they appear to be ones that threaten social status and agency (see also Stroud, Salovey, & Epel, 2002).

While these gender differences in stress provocation appear to emerge most clearly after childhood (see Stroud et al., 2009), the present results are consistent with the adult findings. That is, larger cortisol increases were observed for the girls who, in addition to being anxious and vigilant, were also less socially integrated into the day-care setting. Our scoring of positive social integration was based on whether the child was central to the action, such that if the child left the activity or decided to change the activity then the other children would follow or play would stop. When combined with anxious, vigilant behavior, our summary score may have reflected anxiety over threats to goals of having and maintaining relationships with peers. If by this young age social relationships are beginning to emerge as a more central goal for girls than boys, then this might explain its greater association with the cortisol rise for girls at day care. Note that this argument is consistent with evidence that by preschool age, girls use somewhat more relational aggression or actions that threaten relationships than do boys (Crick et al., 1997).

This argument does not explain why we obtained a significant association between angry, aggressive behavior and the rise in cortisol for boys but not girls. As noted, our finding for boys was consistent with the meta-analysis of aggressive, undercontrolled behavior that provided evidence that while such behavior is associated with low cortisol for school-aged children, among preschool-aged children it is associated with higher cortisol levels (Alink et al., 2008). What was noteworthy was the gender difference in the association of care quality with angry, aggressive behavior. For both sexes, this behavior was observed more when children were receiving less warm and supportive care, but only for boys was angry, aggressive behavior associated with higher levels of intrusive, overcontrolling care. In fact, as suggested in Figure 2, it was at high levels of intrusive, overcontrolling care that the gender differences in angry, aggressive behavior were most pronounced. This suggests that not only the frequency but the social meaning and function of angry, aggressive behavior may have differed for boys and girls. For boys it may have been a reflection of salient threats to agency, which, even at this age, may be a more central goal for boys than girls. Indeed, this would be consistent with evidence that as early as 1 year of age, loss of control (agency) over a loud, noise-making toy elicits more negative affectivity in boys than girls (e.g., Gunnar, 1980), as well as evidence of gendered differences in adulthood in the role of fighting and asserting dominance or control versus seeking relationships in coping with threatening situations (Taylor et al., 2000).

The data for angry, aggressive behavior and stress for boys, however, were even more nuanced than gendered association with intrusive, overcontrolling care would suggest. Specifically, in the moderation analyses we found that among boys, angry and aggressive behavior moderated the association between warm, supportive care and the cortisol rise. Specifically, we found that it was the boys who scored low on this measure whose cortisol rise at child care was associated with the degree of warm and supportive care they received. Under conditions of low warmth and support they showed large increases in cortisol over the day, while under conditions of high warmth they show small increases that were well within the norm for cortisol changes over the day at home. Boys scoring high on angry, aggressive behavior did not show any modulation of the cortisol response to child care in relation to care provider warmth and support. This moderation effect was the opposite of that predicted based on arguments that children who are more vulnerable to behavior problems (e.g., externalizing-type problems in this case) might be more sensitive to variations in care quality. However, as boys were more aggressive than girls, it may be that boys scoring low on angry, aggressive behavior may have had more difficulty managing play with other boys and thus may have been more dependent on the degree of warm, supportive care they were receiving from the care provider.

We recognize that our explanations for the gender differences in cortisol-behavior associations are speculative and that these findings will contribute to continued uncertainty about how behavior in peer settings is associated with activity of the HPA axis in young children. However, they should also encourage researchers to consider the role of gender differences in future studies of child-care stress. Our findings suggest that even when boys and girls do not differ in their physiological stress reactions to complex social situations, they may differ in how emotional-behavioral dispositions and patterns of action relate to their stress responses in those contexts. Furthermore, it seems likely given our pattern of results that these gender differences reflect the gendered aspects of children's social worlds and socialization experiences. As noted above, by adulthood there is good evidence for gender differences in stress reactivity and regulation (Dedovic et al., 2009; Taylor et al., 2000); thus, an increased focus on gendered patterns of stress-behavior-context associations during development is needed.

One of the challenges of interpreting the present findings is that the associations we found accounted for only modest amounts of the variance in the cortisol rise. For example, our measure of intrusive, overcontrolling behavior only accounted for 4% of the variance in the cortisol rise, while in combination the child behaviors and their interaction with sex explained 10% of the variance. It is likely that our methods limited the amount of variance that we could explain. First, cortisol and behavior were measured on different days, and there was likely error in our measures such that behavior on one the days of observation was not exactly the behaviors we would have observed had we measured on the days of cortisol assessment. Unfortunately, the M-ORCE, like the ORCE, requires summating over the 2 days of observation; thus, we do not have a measure of day-to-day variability in either our two process measures of quality or our two measures of child behavior. For the cortisol rise, although we observed a significant correlation across the 2 days of assessment, the rise on day one explained only 16% of the variance in the rise on Day 2. Thus, limitations in the method likely limited the magnitude of the associations we could hope to find. Nonetheless, the effect sizes we observed were consistent with others often reported in child-care studies, and it has been argued that such effect sizes should not be dismissed in considering the import of findings for child-care policy and practice (McCartney & Rosenthal, 2000).

Our findings provide added impetus to the effort to understand whether and how HPA stress responses at child-care impact children's development. Although early life stress is typically viewed as a risk factor for healthy development, recent studies in both nonhuman primates and in human populations raise questions about this type of blanket assessment of early life stressors. For example, Lyons and Parker (2007) have examined the impact of repeated, 1-hr separations in infant squirrel monkeys. As in the work on cortisol responses to full-day child care, these separations in squirrel monkey infants produced marked and repeated activations of the HPA axis. However, followed into the late juvenile and early adult age, animals exposed to this form of early life stress were found to be less fearful, to produce lower rather than higher cortisol responses to stressors, and to show more optimal development of prefrontal regulatory brain circuits; consistent with these findings, they also performed better on tests of executive functioning. Thus, at least for this animal model, repeated separation stress early in life fostered a form of resilience.

In studies of human populations, most of the work on early maltreatment and risky family patterns has been conducted on adults with various affective disorders. Here, the evidence is that early life stress produces hyper-reactivity of the HPA axis, consistent with views that early life stress enhances the neural substrate underlying vulnerability to affective disorders (e.g., Heim, Plotsky, & Nemeroff, 2004). However, in recent studies, researchers have sought to examine the associations between early life stressors and HPA reactivity among adults who are free of significant internalizing pathology. These studies are showing hypoactivity of the HPA axis compared to healthy adults without early life stressors (Carpenter et al., 2007; Elzinga et al., 2008). These findings are also consistent with recently reported evidence that teenagers who experienced more full-time, center-based day care under the age of three produce lower than average levels of cortisol early in the morning at the peak of the diurnal cycle (Roisman et al., 2009). We clearly need to continue to be cautious about interpreting the implications of cortisol elevations in day care. Nonetheless, evidence that this rise is associated with anxious, vigilance (i.e., internalizing) behavior for girls and anger, aggression (i.e., externalizing) behavior for boys should not make us sanguine about its potential implications for children's development.

Several limitations in this study also should be noted. First, although we included family demographic factors in our analyses, we did not examine the quality of parent–child relations and their associations with the day-care cortisol rise. As in other studies of child-care effects, it is possible that issues in the children's families carry more of the weight in predicting how the child responds to child care than issues within the child-care setting (Ahnert & Lamb, 2003; Love et al., 2003; Phillips et al., 2006). Although we did not attempt to pre-select for parents with fairly high educations, incomes, and marital status, the variation in family characteristics was quite narrow, reflecting fairly advantaged and low-risk segments of the U.S. population. It may well be the case, as has been found in prior research (Loeb, Fuller, Kagan, & Carrol, 2004; NICHD ECCRN, 2001, 2003), that quality of child care matters more for low-income children and may thus play a stronger role as an influence on cortisol elevations for children from disadvantaged backgrounds. We were not able to randomly assign children to day-care homes, and thus selection factors could not be controlled in our analyses. We only included licensed day-care homes in our analyses, and thus cannot generalize to effects in the many unlicensed settings used for day care. Finally, although we worked to increase our racial-ethnic variation, the sample was still predominantly White and of European heritage. This again limits generalizability.

With these limitations, however, we have clearly documented that, similar to children in center-based care, many children in family day-care settings show a substantial rise in cortisol over the day, and they do not show this rise when at home. We have found that this rise is observed even in care settings with only two or three children, thus arguing against the hypothesis that social complexity is a major factor underlying increasing activation of the HPA stress response. We also noted associations between intrusive, overcontrolling care and the rise in cortisol for both boys and girls and between warm, supportive care and the cortisol rise for less angry, aggressive boys. These findings may help refine our attempts to understand aspects of child care that promote increases in cortisol in young children. Finally, the gendered differences in associations we noted should focus attention on understanding how boys and girls may differentially process and experience stressors in child-care settings.

References

Ahnert, L., & Lamb, M. E. (2003). Shared care: Establishing a balance between home and child care settings. *Child Development*, *74*, 1044–1049.

Alink, L. R. A., van IJzendoorn, M. H., Bakermans-Kranenburg, M. J., Mesman, J., Juffer, F., & Koot, H. M. (2008). Cortisol and externalizing behavior in children and adolescents: Mixed meta-analytic evidence for the inverse relation of basal cortisol and cortisol reactivity with externalizing behavior. *Developmental Psychobiology*, *50*, 427–450.

Barber, B. K., Stolz, H. E., & Olsen, J. A. (2005). Parental support, psychological control, and behavioral control: Assessing relevance across time, culture, and method. *Monographs of the Society for Research in Child Development*, *70*, 1–137.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182.

Belsky, J., Hsieh, K. H., & Crnic, K. (1998). Mothering, fathering, and infant negativity as antecedents of boys' externalizing problems and inhibition at age 3 years: Differential susceptibility to rearing experience? *Development and Psychopathology*, *10*, 301–319.

Belsky, J., Vandell, D. L., Burchinal, M., Clarke-Stewart, K. A., McCartney, K., Owen, M. T., et al. (2007). Are there long-term effects of early child care? *Child Development*, *78*, 681–701.

Boyce, W. T., & Ellis, B. J. (2005). Biological sensitivity to context: I. An evolutionary developmental theory of the origins and functions of stress reactivity. *Development and Psychopathology*, *17*, 271–301.

Card, N. A., Stucky, B. D., Sawalani, G. M., & Little, T. D. (2008). Direct and indirect aggression during childhood and adolescence: A meta-analytic review of gender differences, inter correlations, and relations to maladjustment. *Child Development*, *79*, 1185–1229.

Carpenter, L. L., Carvalho, J. P., Tyrka, A. R., Wier, L. M., Mello, A. F., Mello, M. F., et al. (2007). Decreased adrenocorticotropic hormone and cortisol responses to stress in health adults reporting significant childhood maltreatment. *Biological Psychiatry*, *62*, 1080–1087.

Clements, A. D., & Parker, R. C. (1998). The relationship between salivary cortisol concentrations in frozen versus mailed samples. *Psychoneuroendocrinology*, *23*, 613–616.

Crick, N. R., Casas, J. F., & Mosher, M. (1997). Relational and overt aggression in preschool. *Developmental Psychology*, *33*, 570–588.

Crick, N. R., Ostrov, J. M., Burr, J. E., Cullerton-Sen, C., Jansen-Yeh, E., & Ralston, P. (2006). A longitudinal study of relational and physical aggression in preschool. *Journal of Applied Developmental Psychology*, *27*, 254–268.

Dedovic, K., Wadiwalla, M., Engert, V., & Pruessner, J. C. (2009). The role of sex and gender socialization in stress reactivity. *Developmental Psychology*, *45*, 45–55.

Dettling, A. C., Gunnar, M. R., & Donzella, B. (1999). Cortisol levels of young children in full-day childcare centers: Relations with age and temperament. *Psychoneuroendocrinology*, *24*, 519–536.

Dettling, A., Parker, S. W., Lane, S., Sebanc, A., & Gunnar, M. R. (2000). Quality of care and temperament determine whether cortisol levels rise over the day for children in full-day child care. *Psychoneuroendocrinology*, *25*, 819–836.

Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin*, *130*, 355–391.

Elzinga, B. M., Roelofs, K., Tollenaar, M. S., Bakvis, P., van Pelt, J., & Spinhoven, P. (2008). Diminished cortisol response to psychosocial stress associated with lifetime adverse events: A study among healthy young subjects. *Psychoneuroendocrinology*, *33*, 227–237.

Fabes, R. A., Martin, C. L., & Hanish, L. D. (2003). Young children's play qualities in same-, other-, and mixed-sex peer groups. *Child Development*, *74*, 921–932.

Geoffroy, M., Côté, S. M., Parent, S., & Séguin, J. R. (2006). Daycare attendance, stress, and mental health. *Canadian Journal of Psychiatry*, *51*, 607–615.

van Goozen, S. H., Fairchild, G., Snoek, H., & Harold, G. T. (2007). The evidence for neurobiological model of childhood antisocial behavior. *Psychological Bulletin*, *133*, 149–182.

Gunnar, M. R. (1980). Control, warning signals and distress in infancy. *Developmental Psychology*, *16*, 281–289.

Gunnar, M. R., & Donzella, B. (2002). Social regulation of the cortisol levels in early human development. *Psychoneuroendocrinology*, *27*, 199–220.

Gunnar, M. R., & Quevedo, K. (2007). The neurobiology of stress and development. *Annual Review of Psychology*, *58*, 145–173.

Gunnar, M. R., Sebanc, A. M., Tout, K., Donzella, B., & van Dulmen, M. M. H. (2003). Temperament, peer relationships, and cortisol activity in preschoolers. *Developmental Psychobiology*, *43*, 346–358.

Gunnar, M. R., Tout, K., de Haan, M., Pierce, S., & Stansbury, K. (1997). Temperament, social competence, and adrenocortical activity in preschoolers. *Developmental Psychobiology*, *31*, 65–85.

Heim, C., Plotsky, P., & Nemeroff, C. B. (2004). The importance of studying the contributions of early adverse experiences to the neurobiological findings in depression. *Neuropsyhopharmacoclogy*, *29*, 641–648.

Howes, C., Galinsky, E., Shinn, M., Gulcur, L., Clements, M., Sibley, A., et al. (1996). *The Florida Child Care Quality Improvement Study*. New York: Families and Work Institute.

Hughes, C., & Dunn, J. (2007). Children's relationships with other children. In C. A. Brownell & C. B. Kopp (Eds.), *Socioemotional development in the toddler years: Transitions and transformations* (pp. 177–200). New York: Guilford.

Kagan, J., Reznick, J. S., & Snidman, N. (1987). The physiology and psychology of behavioral inhibition in children. *Child Development*, *58*, 1459–1473.

Kirschbaum, C., Pirke, K. M., & Hellhammer, D. (1993). The "Trier Social Stress Test": A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, *28*, 76–81.

Kirschbaum, C., Pruessner, J. C., Stone, A. A., Federenko, I., Gaab, J., Lintz, D., et al. (1995). Persistent high cortisol responses to repeated pyschological stress in a subpopulation of healthy men. *Psychosomatic Medicine*, *57*, 468–474.

Kryzer, E. M., Kovan, N., Phillips, D. A., Domagall, L., & Gunnar, M. R. (2007). Toddlers' and preschoolers' experience in family daycare: Age differences and behavioral correlates. *Early Childhood Research Quarterly*, *22*, 451–466.

Kudielka, B., Broderick, J. E., & Kirschbaum, C. (2003). Compliance with saliva sampling protocols: Electronic monitoring reveals invalid cortisol daytime profiles in noncompliant subjects. *Psychosomatic Medicine*, *65*, 313–319.

Legendre, A. (2003). Environmental features influencing toddlers' bioemotional reactions in daycare centers. *Environment and Behavior*, *35*, 523–549.

Lewis, C., Watson, M., & Schaps, E. (2003). Building community in school: The Child Development Project. In M. J. Elias, H. Arnold, & C. S. Hussey (Eds.), *EQ + IQ = Best leadership practices for caring and successful schools* (pp. 100–108). Thousand Oaks, CA: Corwin Press.

Loeb, S., Fuller, B., Kagan, S. L., & Carrol, B. (2004). Child care in poor communities, early learning effects of type, quality, and stability. *Child Development*, *75*, 47–65.

Love, J. M., Harrison, L., Sagi-Schwarz, A., van IJzendoorn, M. H., Ross, C., Ungerer, J. A., et al. (2003). Child care quality matters: How conclusions may vary with context. *Child Development*, *74*, 1021–1033.

Lyons, D. M., & Parker, K. J. (2007). Stress inoculation-induced indications of resilience in monkeys. *Journal of Traumatic Stress*, *20*, 423–433.

McCartney, K., & Rosenthal, R. (2000). Effect size, practical importance, and social policy for children. *Child Development*, *71*, 173–180.

National Institute of Child Health and Human Development Early Child Care Research Network. (1998). Early daycare and self-control, compliance, and problem behavior at twenty-four and thirty-six months. *Child Development*, *69*, 1145–1170.

National Institute of Child Health and Human Development Early Child Care Research Network. (1999). Child outcomes when daycare center classes meet recommended standards for quality. *American Journal of Public Health*, *89*, 1072–1077.

National Institute of Child Health and Human Development Early Child Care Research Network. (2000). Characteristics and quality of daycare for toddlers and preschoolers. *Applied Developmental Science*, *4*, 116–135.

National Institute of Child Health and Human Development Early Child Care Research Network. (2001). Child care and children's peer interaction at 24 and 36 months: The NICHD study of early daycare. *Child Development*, *72*, 1478–1500.

National Institute of Child Health and Human Development Early Child Care Research Network. (2003). Does amount of time spent in child care predict socioemotional adjustment during the transition to kindergarten? *Child Development*, *74*, 976–1005.

National Institute of Child Health and Human Development Early Child Care Research Network. (2010). Testing a series of causal propositions relating time spent in child care to children's externalizing behavior. *Developmental Psychology*, *46*, 1–17.

National Institute of Child Health and Human Development Study of Early Child Care and Youth Development. (2008). Unpublished raw data.

Phillips, D., Crowell, C., Sussman, A., Hane, A., Fox, N. A., & Gunnar, M. R. (2010). *Child care and children's temperament: A story of moderation.* Manuscript submitted for publication.

Phillips, D., McCartney, K., & Sussman, A. (2006). Daycare and early development. In K. McCartney & D. Phillips (Eds.), *Handbook of early childhood development* (pp. 471–489). Malden, MA: Blackwell.

Roisman, G. I., Susman, E., Barnett-Walker, K., Booth-LaForce, C., Owen, M. T., Belsky, J., et al. (2009). Early family and child-care antecedents of awakening cortisol levels in adolescence. *Child Development*, *80*, 907–920.

Rosen, J. B., & Schulkin, J. (1998). From normal fear to pathological anxiety. *Psychological Review*, *105*, 325–350.

SAS Institute. (2004). *SAS 9.1 user's guide.* Cary, NC: Author.

Sebanc, A. M., Pierce, S. L., Cheatham, C. L., & Gunnar, M. R. (2003). Gendered social worlds in preschool: Dominance, peer acceptance, and assertive social skills in boys' and girls' peer groups. *Social Development*, *12*, 91–106.

Sims, M., Guilfoyle, A., & Parry, T. S. (2006). Children's cortisol levels and quality of child care provision. *Child: Care, Health and Development*, *32*, 453–466.

Stroud, L. R., Foster, E., Handwerger, K., Papadonatos, G. D., Granger, D., Kivlighan, K. T., et al. (2009). Stress response and the adolescent transition: Performance versus peer rejection stress. *Development and Psychopathology*, *21*, 47–68.

Stroud, L. R., Salovey, P., & Epel, E. S. (2002). Sex differences in stress responses: Social rejection versus achievement stress. *Biological Psychiatry*, *52*, 318–327.

Talge, N. M., Donzella, B., Kryzer, E., Gierens, A., & Gunnar, M. R. (2005). It's not that bad: Error introduced by oral stimulants in salivary cortisol research. *Developmental Psychobiology*, *47*, 369–376.

Taylor, S. E., Klein, L., Lewis, B. P., Gruenewald, T. L., Gurung, R. A., & Updegraff, J. A. (2000). Biobehavioral responses to stress in females: Tend-and-befriend, not fight-or-flight. *Psychological Review*, *1007*, 411–429.

Tout, K., de Haan, M., Campbell, E. K., & Gunnar, M. R. (1998). Social behavior correlates of cortisol activity in child care: Gender differences and time-of-day effects. *Child Development*, *69*, 1247–1262.

Van Cauter, E., & Refetoff, S. (1985). Evidence for two subtypes of Cushing's disease based on the analysis of episodic cortisol secretion. *New England Journal of Medicine*, *312*, 1343–1346.

Vermeer, H. J., & van IJzendoorn, M. H. (2006). Children's elevated cortisol levels at daycare: A review and meta-analysis. *Early Childhood Research Quarterly*, *21*, 390–401.

Votruba-Drzal, E., Coley, R. L., & Chase-Lansdale, P. L. (2004). Child care and low-income children's development: Direct and moderated effects. *Child Development*, *75*, 296–312.

Watamura, S. E., Donzella, B., Alwin, J., & Gunnar, M. R. (2003). Morning to afternoon increases in cortisol concentrations for infants and toddlers at child care: Age differences and behavioral correlates. *Child Development*, *74*, 1006–1020.

Watamura, S. E., Donzella, B., Kertes, D. A., & Gunnar, M. R. (2004). Developmental changes in baseline cortisol activity in early childhood: Relations with napping and effortful control. *Developmental Psychobiology*, *45*, 125–133.

Watamura, S. E., Sebanc, A. M., & Gunnar, M. R. (2002). Rising cortisol at childcare: Relations with nap, rest, and temperament. *Developmental Psychobiology*, *40*, 33–42.

Child Development, May/June 2010, Volume 81, Number 3, Pages 870–883

Children's Sleep and Adjustment Over Time: The Role of Socioeconomic Context

Mona El-Sheikh, Ryan J. Kelly, Joseph A. Buckhalt, and J. Benjamin Hinnant
Auburn University

Relations were examined between children's sleep and their externalizing and internalizing symptoms. Longitudinal relations were examined when children were in 3rd (T1) and 5th (T2) grades, and cross-sectional relations were assessed at T2. Participants included 176 children at T1 (M = 8.68 years) and 141 children at T2 (M = 10.70 years). Sleep was examined via subjective reports and actigraphy. Children reported on anxiety, self-esteem, and depression symptoms, and parents reported on children's externalizing and internalizing symptoms. Cross-sectionally and longitudinally, sleep problems were associated with worse adjustment outcomes; African American children or those from lower socioeconomic status homes were at particular risk. Findings highlight the importance of adequate sleep for children's optimal development, especially in the context of ecological risk.

The significance of children's sleep to their developmental outcomes is receiving increased attention in the child development literature. Rapidly developing evidence links children's sleep problems within the normative range and their adjustment outcomes. Behavioral problems (Paavonen et al., 2002) and internalizing symptoms (Gregory et al., 2005) have been associated with shorter sleep duration and worse sleep quality in elementary-school-aged children. Nevertheless, many open scientific questions remain regarding relations between sleep disruptions and child developmental outcomes. Notably, few studies have investigated these associations with objective sleep measures, samples of typically developing children, multiple informants of adjustment outcomes, multiple response domains, or longitudinal designs. The main objective of this study is to assess relations between sleep disruptions (amount, quality, schedule) and children's adjustment both cross-sectionally at T2 and longitudinally with two waves of data. This investigation builds on the existing literature by examining sleep problems both objectively via actigraphy and subjectively via child and parent reports in a large community sample utilizing a longitudinal design. Mothers', fathers', and children's reports of child adjustment are used for a better explication of child functioning in the context of sleep disruptions. Further, we explore the role of race, socioeconomic status (SES), and gender in moderating relations between sleep and adjustment outcomes.

For all children, acquisition of adaptive socioemotional behaviors is a primary developmental task with immediate and long-term consequences. Sleep has only recently been related to a number of outcomes in this domain, and indications are that sleep plays a vital role in healthy development. In community samples, restricted or disrupted sleep has been associated with externalizing behaviors, including aggression, conduct problems, and delinquency (Aronen, Paavonen, Fjallberg, Soininen, & Torronen, 2000; Chervin, Dillon, Archbold, & Ruzicka, 2003). Similarly, sleep problems in childhood have been associated with internalizing problems, primarily symptoms of anxiety and depression (Chorney, Detweiler, Morris, & Kuhn, 2008; Gregory & Eley, 2005), which may persist to adolescence and adulthood (Gregory & O'Connor, 2002; Gregory et al., 2005). Theory to guide research in this domain is in an emergent stage, but Dahl (1996) hypothesizes a role for the prefrontal cortex (PFC) in the association of sleep problems and emotional dysregulation. Sleep deprivation in adults is known to disrupt behaviors mediated in the PFC,

---

This research was supported by National Science Foundation Grants 0339115 and 0623936, and Alabama Agricultural Experiment Station/Lindsey Foundation Grant ALA080-001. We acknowledge contributions made by staff of our Research Laboratory, most notably Lori Staton and Bridget Wingo, for data collection. We also thank school personnel, and children and parents who participated in this study.

Correspondence concerning this article should be addressed to Mona El-Sheikh, Human Development and Family Studies, 203 Spidle Hall, Auburn University, Auburn, AL 36849-5214. Electronic mail may be sent to elshemm@auburn.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8103-0015

including executive processes necessary for controlling cognition and emotion (Muzur, Pace-Schott, & Hobson, 2002).

The term *sleep problems* as used here does not refer to clinically diagnosed sleep disorders (e.g., sleep apnea). Sleep problems were examined in typically developing children, are conceptualized along a continuum, and are indicated by shorter average sleep times, sleep characterized by higher activity or several extended night wakings, or sleep schedules that are more variable from night to night, relative to other participants. Arguably, chronic sleep insufficiency and erratic sleep schedules may be the most prevalent forms of sleep problems in otherwise typically developing children, and assessment of sleep problems along the normative continuum is critical for clarifying developmental outcomes for many children in the population.

With the present sample, cross-sectional data collected in the first wave, when children were in third grade, supported relations between sleep problems and children's externalizing and internalizing symptoms (El-Sheikh, Buckhalt, Cummings, & Keller, 2007). This study reports cross-sectional findings from the second wave of data when children were in fifth grade, as well as longitudinal data over the 2 years of assessment. Multiple child outcomes (externalizing and internalizing symptoms) were examined.

One study objective is the assessment of child race and SES as moderators of the link between sleep problems and child adjustment. Research with adults indicates higher rates of clinical sleep disorders (for a review, see Durrence & Lichstein, 2006) and more restricted sleep (Stamatakis, Kaplan, & Roberts, 2007) in African Americans (AAs) in comparison with European Americans (EAs). Fewer studies have examined race- and SES-related effects in children's sleep, and in the majority that did, clinical samples were used. For example, evidence indicates that AA children have higher rates of asthma and sleep-disordered breathing, less-reported sleep duration, and poorer sleep quality (Crosby, LeBourgeois, & Harsh, 2005; Rosen et al., 2003). Further, in comparison with EA children, AA children are more likely to have later bedtimes, earlier wake times (Adam, Snell, & Pendry, 2007), less consistent bedtime routines, and to cosleep with siblings, even after controlling for SES factors (Milan, Snow, & Belay, 2007). In our own research with the first wave of data obtained with the present sample, controlling for SES, EA children slept longer, had lower levels of disrupted sleep, and reported fewer sleep problems in comparison with AA children (El-Sheikh, Buckhalt, Cummings, et al., 2007; El-Sheikh, Buckhalt, Keller, Cummings, & Acebo, 2007). Additionally, controlling for race, children from higher SES backgrounds slept longer and had fewer sleep problems than children from lower SES backgrounds (El-Sheikh, Buckhalt, Keller, Cummings, et al., 2007).

Although examining differences in sleep parameters between AA and EA children is informative, explication of developmental outcomes for children in various cultural contexts requires the testing of the racial-cultural milieu as a moderator of effects. Very few studies have undertaken such assessments in the context of child sleep and development. One objective here is to examine the role of race (AA, EA) and family SES as moderators of the link between sleep problems and child adjustment. In two of the previous studies with the present sample at T1, we found that race and SES moderated relations between sleep and cognitive and achievement tasks (Buckhalt, El-Sheikh, & Keller, 2007; El-Sheikh, Buckhalt, Keller, Cummings, 2007). In a recent follow-up of children in this sample, longitudinal relations between children's sleep and cognitive functioning were examined (Buckhalt, El-Sheikh, Keller, & Kelly, 2009). Disrupted sleep was associated with poorer cognitive performance 2 years later primarily for AA children and those with less educated fathers. These findings are supportive of the health disparities view (e.g., Carter-Pokras & Baquet, 2002) in that minorities and individuals of lower SES are likely exposed to chronic stress and have an increased burden of many adverse conditions. Under adverse conditions, disruptions of a primary biological regulation system like sleep may have a greater deleterious effect. However, we are not aware of any studies that investigated SES, ethnicity, or race as moderators of the link between children's sleep and their socioemotional functioning cross-sectionally or longitudinally.

Several studies have reported gender-related sleep differences among school-aged children; however, findings have been inconsistent. Some studies found that girls had better sleep quality (Buckhalt et al., 2007; El-Sheikh, Buckhalt, Mize, & Acebo, 2006; Sadeh, Raviv, & Gruber, 2000) or longer sleep durations (El-Sheikh, Buckhalt, Cummings, et al., 2007; Fredricksen, Rhodes, Reddy, & Way, 2004). Other research has reported that boys sleep more than girls (Gau & Soong, 1995), and as girls near early adolescence, sleep–wake times become significantly earlier than boys' wake times (Lee, McEnany, & Weekes, 1999). However, there have been no published studies of gender as a moderator of effects in the sleep–child development link. Further, few of the studies that have documented gender-related mean differences in sleep parameters were based on actigraphy, which warrants further investigation to determine gender-related effects in boys' and girls' sleep patterns and their effects on developmental outcomes (Sadeh et al., 2000). We examined gender as a moderator of effects in the association between sleep and child adjustment.

To summarize, the main objective was to examine relations between sleep disruptions (amount, quality, schedule) and children's externalizing and internalizing symptoms cross-sectionally at T2 and longitudinally (T1 and T2). At T1, children were in third grade; at T2 they were in the fifth grade. Building on T1 cross-sectional evidence (El-Sheikh, Buckhalt, Cummings, et al., 2007), we expected that reduced sleep amount, less-optimal sleep quality, and an inconsistent sleep schedule would be associated with increased adjustment problems at T2. Longitudinally, we expected that T1 sleep problems would be associated with more adjustment problems, after controlling for autoregressive effects. Thus, longitudinal analyses examined sleep as a predictor of *changes* in adjustment outcomes over time, controlling for T1 child adjustment. Multiple objective and subjective sleep measures were examined for a better clarification of relations between sleep and child adaptation (Sadeh, 2008).

A second objective was to examine moderators of effects in the link between sleep disruptions and child adjustment: SES, race, and gender. Based on prior empirical evidence (El-Sheikh, Buckhalt, Cummings, et al., 2007) and the health disparities view (Carter-Pokras & Baquet, 2002), we anticipated that the link between sleep disruptions and child adjustment would be more robust for either AA children or children from lower SES homes. SES and race were examined as cross-sectional (T2) and longitudinal moderators. Given the lack of conceptual or empirical guidance in the literature, we had no specific hypotheses regarding the role of gender in the sleep–adjustment link.

Method

Participants

Children in third grade were recruited from public schools in the Southeastern United States. Parent contact information was obtained from the schools. To participate, children had to satisfy the inclusion criteria, which included living in a two-parent home, and having no history or diagnosis of a sleep problem, attention deficit hyperactivity disorder, chronic or acute physical illness, mental retardation, or a learning disability. T1 participants included 176 children (78 boys, 98 girls) with an average age of 8.68 years ($SD = 0.36$), along with their parents. Mothers' reports on the Puberty Development Scale (1 = prepubertal, 2 = early pubertal, 3 = midpubertal, 4 = late pubertal, 5 = postpubertal; Petersen, Crockett, Richards, & Boxer, 1988) indicated that the majority of boys ($M = 1.25$, $SD = 0.23$) and girls ($M = 1.45$, $SD = 0.34$) were prepubertal.

For SES classifications, families were categorized based on Hollingshead's (1975) criteria. Specifically, 28% were classified as Level 1 or 2 (e.g., unskilled workers), 41% were in Level 3 (e.g., skilled workers), and 31% were in Level 4 or 5 (e.g., minor professional or professional).

Racial identity of the child was identified by the parent according to 2000 U.S. census categories. All parents selected a sole category of "White" or "Black or AA" with the following result: 69% EA, 31% AA. The sample constitution is very similar to the racial and SES composition of the school district from which it was drawn, which includes 75% White and 22% Black or AA. We oversampled to include a wide range of SES from both AA and EA families with a resulting low but still significant association between race and SES ($r = -.24$, $p < .01$ for SES raw score; $r = -.19$, $p < .05$ for SES level; note that race coding used 1 for AAs and 0 for EAs). In this study, the term *race* rather than *ethnicity* is used. Much attention has been given to the proper use and measurement of race and ethnicity in scientific research (e.g., Entwisle & Astone, 1994), and we elected to use race.

Approximately 2 years later, 142 of the families returned for a second wave of data collection (T2; 81% of the original sample). Attrition occurred due to moving out of the study area, inability to be located, disinterest in continuing study participation, and a busy schedule. At T2, 80 boys and 61 girls (70% EA, 30% AA) participated. Children's average age was 10.70 years ($SD = 0.55$) at T2. Average puberty status at T2 was 2.06 ($SD = 0.55$) for girls and 1.49 ($SD = 0.42$) for boys, corresponding with an early pubertal status for girls and prepubertal status for boys (Petersen et al., 1988).

Independent samples t tests and χ^2 analyses were used to determine whether differences existed between children who participated in one or two waves. In comparison with children who participated during both waves, children who did not participate in the second wave came from families with less income, $t(162) = -2.18$, $p < .05$, and lower SES status, $t(174) = -3.46$, $p < .001$. There were no gender, race, age, pubertal stage, externalizing, internalizing, or sleep parameter differences between retained and missing participants, with the exception that children who dropped out were reported by their parents to have more sleep-related difficulties, $t(171) = 2.16$, $p < .05$. No differences in objective sleep measures were found.

Procedure

A research assistant instructed parents to place an actigraph on the child's nondominant wrist just prior to bedtime for seven consecutive nights. To help ensure typical sleep patterns, actigraphic data collection occurred only during the regular school year, excluding holidays, and only data from medication-free nights were used for analyses (for more detail regarding procedures, see Buckhalt et al., 2009; El-Sheikh, Buckhalt, Keller, Cummings, et al., 2007).

Typically, children came to our university laboratory with their mothers and fathers following the last day of data collection. Each parent and the child completed subjective sleep and child adjustment measures. Children were administered each measure via interview by a research assistant, whereas parents used self-report. Child height and weight were measured in the lab to compute body mass index (BMI), based on Must, Dallal, and Dietz's (1991) criteria. The study was approved by the University Institutional Review Board and parents and children provided informed consent and assent. Procedures and measures at T1 and T2 are the same unless otherwise indicated.

Measures

Socioeconomic status. Two different markers were used to determine SES: parents' combined annual family income and SES classifications based on Hollingshead's (1975) criteria. For annual income, families were classified into one of the following categories: (a) \$10,000–\$20,000, (b) \$20,000–\$35,000, (c) \$35,000–\$50,000, (d) \$50,000–\$75,000, or (e) more than \$75,000. Because annual income and Hollingshead SES classifications were correlated at T1 ($r = .46$, $p < .001$) and T2 ($r = .49$, $p < .001$), they were composited at each time point to reduce the number of analyses and to create an overall SES score. Reference to SES in the remainder of this study refers to the composite.

Parent report of children's sleep. Mothers and fathers completed the 35-item Children's Sleep Habits Questionnaire (CSHQ; Owens, Spirito, & McGuinn, 2000), which has good validity and reliability for community and clinical samples. The CSHQ assesses several sleep domains, including bedtime resistance, sleep onset delay, sleep duration, sleep anxiety, parasomnias, night awakenings, sleep disordered breathing, and daytime sleepiness, during the last week. For each item, response options are *zero to one time* (1), *two to four times* (2), and up to *five to seven times* (3). All items were combined to form a total score. Mothers' and fathers' reports were associated ($r = .55$ at T1; $r = .40$ at T2; $ps < .01$), and were averaged to derive a parent reported sleep problems composite (coefficient αs = .90 at T1 and .95 at T2). At T1, 95% of mothers and 81% of fathers completed the CSHQ, and at T2, 94% of mothers and 74% of fathers completed the measure. When one reporter's data were missing, the other's data were used in the analyses. At T1 and T2, scores ranged from 37 to 89, and 35 to 79, respectively.

Child self-reports of sleep. The Sleep Habits Survey (Wolfson & Carskadon, 1998) was used to obtain subjective self-reports of children's sleep, and has been shown to be reliable and valid for children as young as 6 years (e.g., El-Sheikh & Buckhalt, 2005). Two subscales were pertinent: the 10-item Sleep–Wake Problems scale (coefficient αs = .71 at T1 and .79 at T2) and the nine-item Sleepiness scale (coefficient αs = .77 at T1 and .67 at T2). Items on the Sleep–Wake Problems scale assessed irregular sleep times, staying up past bedtime, and oversleeping. For each item, response options ranged from *never* (1) to *everyday/night* (5), with a higher number indicative of increased problems. At T1 and T2, scores on the Sleep–Wake Problems scale ranged from 10 to 34, and 10 to 40, respectively. Items on the Sleepiness scale asked whether participants unintentionally fell asleep during nine daily activities, such as in class, watching TV, or while riding in a car; one item regarding feeling sleepy while driving was not administered. Children answered each item using a 5-point response format that ranged from *never* (1) to *everyday/night* (5), with higher scores, reflecting increased problems. At T1, total scores on the Sleepiness scale ranged from 9 to 27; at T2, scores ranged from 9 to 24.

Objective measurement of sleep. Children wore actigraph bracelets, which were roughly the size and weight of a wristwatch, between bedtime and wake time for seven consecutive nights. Actigraphy has been frequently used as a proxy for sleep duration and quality, and has demonstrated good reliability, especially when worn for several consecutive nights (Acebo et al., 1999; Sadeh, Gruber, & Raviv, 2002). To help ensure reliability among actigraphic recordings, parents completed daily sleep logs to cross-check child bedtime and wake times. Actigraphs were Octagonal Basic Motionloggers (Ambulatory Monitoring Inc., Ardsley, NY) and measured body motion in 1-min epochs using the zero crossing mode. Children were considered to be awake or asleep based on Sadeh's scoring algorithm (Sadeh, Sharkey, & Carskadon, 1994). The actigraph and analysis software have demonstrated reliability and validity when compared with polysomnography (Sadeh et al., 1994). Actigraphy was used to derive the following sleep parameters: (a) sleep duration—the total amount of time between sleep onset and wake time, and (b) sleep efficiency—percentage of motionless sleep and generally accepted in the sleep literature as an index of sleep quality. Further, to capture the variability in sleep schedule over the week of actigraphic assessment, variability in sleep onset and variability in wake time were computed, using the mean-centered coefficient of variance statistic (Snedecor & Cochran, 1967), and were composited to create (c) variability in sleep schedule.

For sleep duration and sleep efficiency, scores were generated by computing an average score across all available nights. Intraclass correlations indicated that each sleep variable exceeded Acebo et al.'s (1999) recommended stability level of .70: sleep duration αs = .78 at T1 and .86 at T2; sleep efficiency αs = .87 at T1 and .92 at T2. These coefficients indicate that children's sleep parameters were stable across the various nights of sleep assessment.

Overall, 75% of children had valid actigraph data across all seven nights at T1. In some instances, actigraph recording did not occur on all nights. Reasons for missing data included the use of allergy medicine, mechanical actigraph problems, and forgetting to wear the actigraph. Consequently, at T1, the actigraph was worn an average of 6.59 nights (SD = 0.85), and 14% of the children had data for six nights, 8% had data for five nights, and 3% had data for four nights or fewer. For T2, 56% of participants had data for all seven nights, 25% had data for six nights, 14% had data for five nights, and 5% had data for four nights or fewer (M = 6.30, SD = 0.95). The rates of valid actigraphy data in this study satisfied Acebo et al.'s (1999) recommendations.

Parent reports of child externalizing and internalizing symptoms. Both mothers and fathers reported on child adjustment using the 280-item Personality Inventory for Children (PIC; Wirt, Lachar, Klinedinst, & Seat, 1990). For this study, two scales were used. The Externalizing scale assesses aggression, impulsivity, disruptive behavior, delinquency, and noncompliance. The Internalizing scale measures depression, anxiety, fear, worry, and psychosomatic problems. The PIC has very good psychometric properties (Wirt et al., 1990). Mother and father reports of child externalizing (r = .77 at T1, and .63 at T2, ps < .001) and internalizing symptoms (r = .73 at T1, and .45 at T2, ps < .001) were correlated, and were averaged to derive an externalizing and an internalizing symptoms composites. For externalizing symptoms, αs = .89 at T1 and .93 at T2; for internalizing symptoms, αs = .80 at T1 and .90 at T2.

Child report of internalizing symptoms. Children completed the 28-item Revised Children's Manifest Anxiety Scale (RCMAS; Reynolds & Richmond, 1978). The RCMAS is frequently used to measure child anxiety and has demonstrated test–retest reliability (Wisniewski, Mulick, Genshaft, & Coury, 1987) and concurrent validity (James, Reynolds, & Dunbar, 1994). For this study, α = .89 at T2. Further, children completed the Self-Perception Profile for Children questionnaire (SPPC; Harter, 1982, 1985), which assesses children's perceived competence and has demonstrated test–retest reliability and validity (Harter, 1982). For this study, children's reports on the 6-item Global Self-Worth subscale were used. At T2, the SPPC α = .84. Children reported on depression symptoms using the Children's Depression Inventory (CDI; Kovacs, 1985), which contains 27 items; one item regarding suicidal ideation was not administered. The CDI has demonstrated good test–retest reliability and discriminant validity (Kovacs, 1985). At T2, CDI α = .77.

Bivariate correlations revealed that children's anxiety was associated with depression (r = .49 at T2, p < .001) and self-esteem scores (r = −.42 at T2, p < .001). Further, self-esteem was associated with depression (r = −.47 at T2, p < .001). To reduce the number of analyses, the children's anxiety, self-esteem, and depression scores were composited to create a total child report of internalizing symptoms score. Self-esteem scores were reverse-coded, and children's anxiety, self-esteem, and depression scores were standardized using z scores prior to compositing each measure. Because the child-report internalizing score is a standardized composite, it is likely not an appropriate measure for our longitudinal analyses (Singer & Willett, 2003). Thus, this measure is only included in cross-sectional analyses at T2.

Results

Analysis Plan

In preliminary analyses, descriptive statistics and correlations among variables were examined. Next, a series of cross-sectional and longitudinal path models were estimated: First, demographic characteristics were entered (child gender, race, age, pubertal status, BMI, and SES). Next, main effects of children's T2 sleep parameters (parent-reported sleep problems, child-reported sleep–wake problems or sleepiness, *or* actigraphic measures of sleep duration, efficiency, or variability in sleep schedule) were entered in the models as predictors of children's T2 adjustment (see Figure 1 for an example of a cross-sectional model). This resulted in two distinct models of prediction: one for subjective reports and one for objective actigraphy measures. Within these models, all relevant sleep variables were entered simultaneously to evaluate their relative contributions to prediction while controlling for other sleep characteristics. To avoid potential shared method bias with the models involving subjective reports, only child-reported subjective sleep measures were used to predict parent-reported child adjustment. Similarly, only parents' reports of child sleep were used to predict child-reported internalizing symptoms. Finally, interactions between children's sleep and the potential moderators (race, SES, and child gender) were added to the models.

To predict residual change, longitudinal path models were estimated in which children's sleep at T1 was examined as a predictor of children's adjustment at T2, while controlling for the respective adjustment measures at T1. Controlling for T1 levels of the predicted variable reduces bias in parameter estimates, permits conclusions regarding change in the predicted variable over time, and provides information about the direction of effects between model variables (Cole & Maxwell, 2003). The same analytic process used to evaluate the cross-sectional data was used to evaluate the longitudinal data (see Figure 2 for an example of a longitudinal model). As noted in the Method section, child-reported internalizing symptoms were not analyzed longitudinally. In order to decrease the likelihood of Type I error, moderation effects were only reported and interpreted if the specific moderator was significant at least 10% of the time. For example, 10% of the possible interactions including SES as a moderator between sleep and

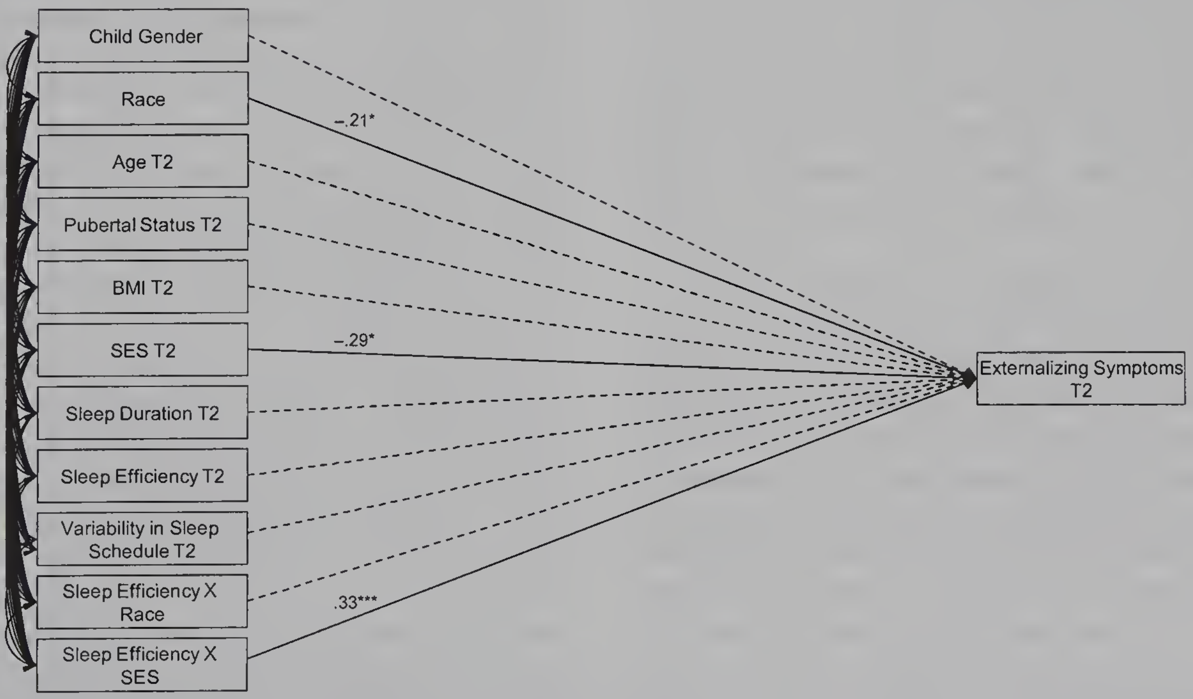


Figure 1. Race and socioeconomic status (SES) as moderators of associations between sleep efficiency at T2 and externalizing symptoms at T2.
Note. Significant pathways are presented with solid lines and nonsignificant pathways are depicted with dotted lines. Race was dummy coded with a 1 for African Americans and a 0 for European Americans.
$^{*}p < .05$. $^{***}p < .001$.

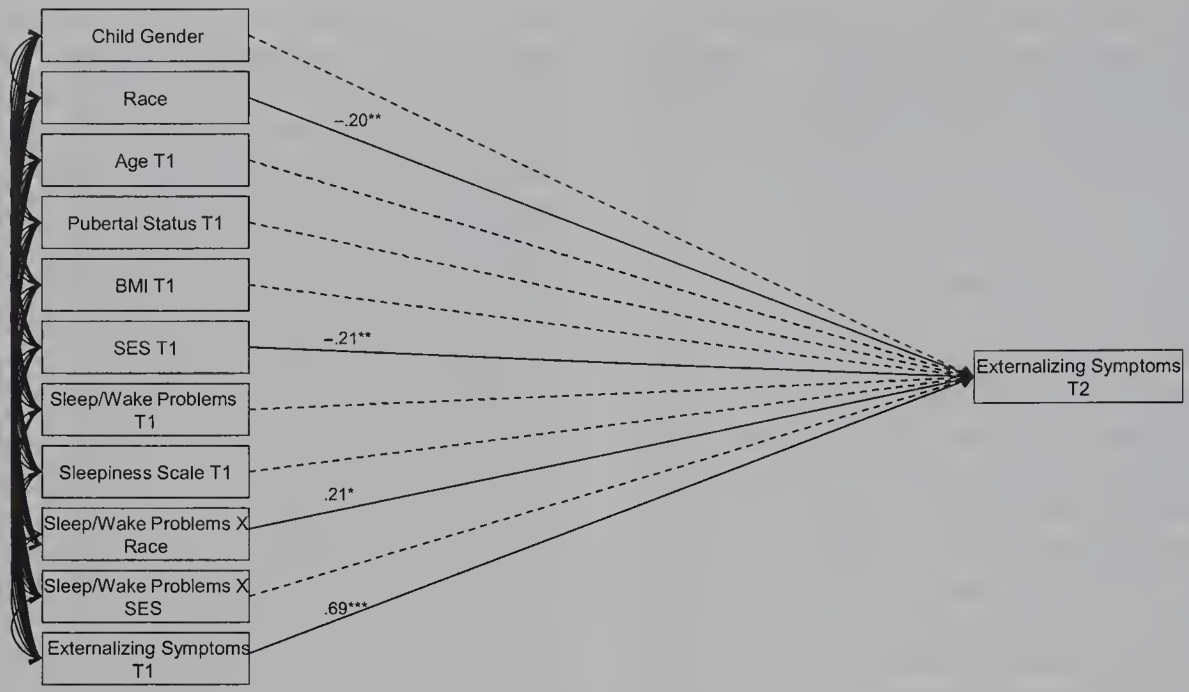


Figure 2. Race and socioeconomic status (SES) as moderators of associations between child-reported sleep–wake problems at T1 and externalizing symptoms at T2.

Note. Significant pathways are presented with solid lines and nonsignificant pathways are depicted with dotted lines. Race was dummy coded with a 1 for African Americans and a 0 for European Americans.

$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

adjustment outcomes would have to be significant in order to be reported.

Path analysis (Amos 16.0; Arbuckle, 2007) was used because of the benefits it has regarding missing data: using maximum likelihood estimation and thus utilizing a larger sample size and more relaxed assumption that data are missing at random than multiple regressions (Little & Rubin, 2002). Exogenous variables were allowed to correlate, resulting in saturated models. Fit indices were not produced because all models were fully saturated. Within each model, all continuous variables were mean-centered.

Path model findings are reported first for cross-sectional analyses and then for longitudinal analyses and are ordered according to the dependent variable as follows: externalizing symptoms, parent-reported internalizing symptoms, and child-reported internalizing symptoms. The prediction of each dependent variable is reported in the following format: (a) demographic characteristics, (b) main effects of subjective or objective sleep, and (c) interactions (if any) between sleep characteristics and moderators. Although gender was tested as a moderator in the link between children's sleep and their adjustment, no statistically significant interactions involving gender were detected for either the cross-sectional or longitudinal associations. Thus, no such moderation effects are reported.

Aiken and West's (1991) criteria were used to interpret interaction effects; Preacher's interaction utility was used to plot significant interactions (Preacher, Curran, & Bauer, 2006). Representative significant interactions were plotted at high (+1 *SD*) and low (−1 *SD*) levels of the predictor and moderator. Slopes representing the association between the predictor and outcome variables were examined to determine if they were significantly different from zero. Of note is that a significant interaction term denotes that the association between the predictor and outcome variables is different for children at higher and lower levels of the moderator (i.e., the plotted lines are significantly different from each other).

Descriptive statistics and associations among variables. Means, standard deviations, and bivariate correlations are provided in Table 1. Children averaged 8 hr 53 min (*SD* = 37 min) of sleep per night at T1 and 8 hr 33 min (*SD* = 38 min) per night at T2. Notably, at T1 and T2, AAs had a shorter sleep duration (*r*s = −.22 and −.24, respectively) and tended to have more subjective sleep-related problems (*r*s = .16–.25). At T1 and T2, SES was positively associated with sleep duration (*r*s = .17 and .19, respectively), such that higher SES children had longer sleep. Several associations between children's sleep and adjustment outcomes were also observed. At both time points, more parent- and

Table 1

Bivariate Correlations and Descriptive Statistics Among Demographic, Sleep, and Adjustment Outcomes

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	*M*	*SD*
1. Child gender (boys = 1, girls = 0)	—																											—	—
2. Race (AA = 1; EA = 0)	−.08	—																										—	—
3. Age T1	−.02	−.10	—																									8.68	.36
4. Puberty T1	−.32*	.29*	.00	—																								1.36	.31
5. BMI T1	−.08	.05	−.06	.09	—																							20.34	4.63
6. SES T1	.09	−.37*	.16*	−.13	−.12	—																						—	—
7. Parent-reported sleep T1	.12	−.05	−.11	.00	−.14	−.16*	—																					48.60	6.99
8. Sleep–wake problems T1	−.05	.25*	−.09	.10	.04	−.13	.07	—																				17.12	4.92
9. Sleepiness scale T1	.15*	.16*	−.10	.05	−.01	−.10	.06	.43*	—																			14.15	4.51
10. Sleep duration T1	−.24*	−.22*	.14	−.08	−.09	.17*	−.01	−.09	−.10	—																		8 hr 53 min	37 min
11. Sleep efficiency T1	−.20*	.09	.17*	−.03	−.17*	.08	−.09	−.06	−.07	.19*	—																	86.56	8.21
12. Variability in sleep schedule T1	.01	.16*	.04	−.02	.12	−.10	.11	−.08	−.08	−.06	.00	—																0.09	0.04
13. Parent-reported externalizing T1	.08	−.04	.10	−.16	−.11	.06	.29*	.02	−.04	−.09	−.25*	.00	—															50.49	10.49
14. Parent-reported internalizing T1	.03	.02	.08	−.11	.04	−.03	.32*	−.03	−.06	−.15	−.30*	.08	.70*	—														51.29	12.46
15. Age T2	−.05	−.09	.75*	−.10	−.06	.15	−.04	.03	−.05	.02	.14	.07	.19*	.24*	—													10.70	0.55
16. Puberty T2	−.53*	.14	.16	.55*	.14	−.05	−.10	.14	.02	.13	.08	−.04	−.12	−.10	.14	—												1.81	0.57
17. BMI T2	−.01	.02	−.04	−.02	.83*	−.12	−.13	.04	−.07	−.10	−.21*	.05	−.05	.03	−.01	.09	—											22.34	6.47
18. SES T2	.04	−.38*	.10	−.14	−.16	.92*	−.10	−.06	−.07	.26*	.14	−.12	−.04	−.12	.14	−.07	−.11	—										—	—
19. Parent-reported sleep T2	.14	−.16	−.05	−.10	−.21*	.00	.58*	.02	.02	−.03	−.11	.02	.38*	.30*	−.07	−.11	−.25*	−.04	—									44.39	5.98
20. Sleep–wake problems T2	.06	.02	−.03	−.01	.08	−.13	.13	.40*	.20*	−.05	−.03	.06	.14	.03	.01	.00	.06	−.22*	.22*	—								15.93	4.67
21. Sleepiness scale T2	.14	.20*	.03	.01	.05	−.03	.12	.20*	.19*	.00	.12	.04	.05	−.02	.00	−.09	.02	−.13	.14	.47*	—							12.61	3.51
22. Sleep duration T2	−.13	−.24*	.00	.09	−.12	.14	.06	−.14	.01	.47*	.11	−.15	−.16	−.34*	−.23*	.04	−.14	.19*	−.05	−.07	.00	—						8 hr 33 min	38 min
23. Sleep efficiency T2	−.13	.03	.04	.04	−.12	.09	−.02	−.02	.02	.09	.41*	−.04	.02	−.02	.08	−.01	−.13	.18*	−.08	−.01	.06	−.02	—					89.15	9.31
24. Variability in sleep schedule T2	.04	.03	.02	.02	.09	−.16	.03	.03	.05	−.15	−.13	.12	.07	.00	.05	.07	.03	−.16	.14	.27*	.26*	−.30*	−.03	—				0.11	0.10
25. Parent-reported externalizing T2	.00	−.13	.01	−.10	−.07	−.11	.43*	−.07	.00	−.06	−.04	.00	.71*	.51*	.11	−.08	−.08	−.19*	.48*	.26*	.18*	−.09	−.03	.11	—			48.26	8.54
26. Parent-reported internalizing T2	.16	−.13	.05	−.19*	−.02	−.03	.44*	−.06	−.06	−.05	−.08	.01	.56*	.65*	.19*	−.16	−.06	−.10	.44*	.24*	.06	−.18*	−.02	.13	.60*	—		48.53	7.99
27. Child-reported internalizing T2	.11	.05	−.10	.06	.09	−.13	.04	.30*	.26*	−.14	−.09	−.02	.16	.12	−.11	−.10	.05	−.17	.16	.52*	.49*	−.15	.00	.34*	.20*	.24*	—	—	—

Note. AA = African American; EA = European American; T1 = data collected at Time 1; T2 = data collected at Time 2; BMI = body mass index; SES = socioeconomic status.
**p* < .05.

child-reported sleep problems were associated with elevated levels of parent- and child-reported externalizing and internalizing problems (*rs* = .18–.52). Regarding objective reports of sleep, more sleep efficiency was associated with less parent-reported externalizing and internalizing symptoms at T1 (*rs* = −.25 and −.30 respectively). At T2, longer sleep duration was negatively correlated with parent-reported internalizing problems (*r* = −.18), and more variability in sleep schedule was positively associated with more child-reported internalizing symptoms (*r* = .34). Also of note, subjective reports of sleep tended to be correlated over time (*rs* = .19–.58), as were sleep duration (*r* = .47) and sleep efficiency (*r* = .41).

Cross-Sectional Prediction of Child Adjustment

Externalizing symptoms. Demographic characteristics accounted for 11% of the variance in externalizing symptoms. Race (β = −.21; *p* = .02) and SES (β = −.29; *p* = .002) were significantly related to externalizing symptoms; AA children were reported as having fewer externalizing symptoms and as SES increased, externalizing symptoms decreased. Child-reported daytime sleepiness, sleep–wake problems, or their interactions were not related to externalizing symptoms, although sleep–wake problems approached significance (β = .16; *p* = .08). No other R^2 is reported because no variable beyond demographics accounted for a significant additional amount of variance in externalizing symptoms.

In predicting externalizing symptoms from objective sleep actigraphy measures, no main effects emerged for sleep duration, efficiency, or variability. One significant interaction was observed between sleep efficiency and SES (β = .33; *p* < .001). For children of lower SES, as sleep efficiency increased, externalizing symptoms decreased significantly (*p* = .01), but for children of higher SES, as sleep efficiency increased, externalizing symptoms increased significantly (*p* = .01; Figure 3A). Importantly, no differences in externalizing symptoms were observed between higher and lower SES children when sleep efficiency was high. However, at lower levels of sleep efficiency, children from higher SES homes had fewer externalizing behaviors than their lower SES counterparts. This interaction accounted for 10% of variance in externalizing symptoms beyond demographic or main effects.

Parent-reported internalizing symptoms. Both child age (β = .22; *p* = .009) and SES (β = −.19; *p* = .03) were associated with parent-reported internalizing symptoms; older children were reported as having more problems, and as SES increased, internalizing symptoms decreased. Demographic characteristics accounted for 11% of variance in internalizing symptoms.

Sleep–wake problems reported by children were positively associated with internalizing symptoms (β = .24; *p* = .009) and accounted for 5% of unique variance. This main effect was qualified by an interaction with SES (β = −.19; *p* = .05) that accounted for 2% of unique variance, which indicated that although sleep–wake problems were unrelated to internalizing symptoms for children from higher SES families, they were positively related to symptoms for children from lower SES

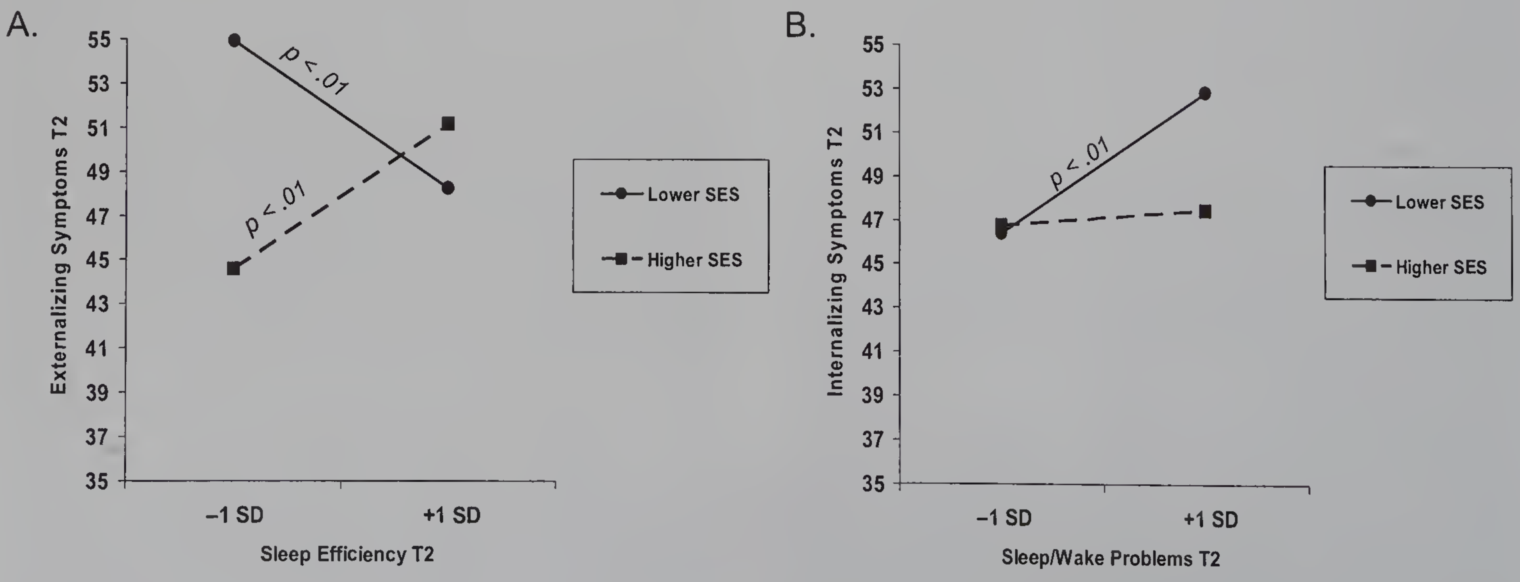


Figure 3. Cross-sectional interactions between socioeconomic status (SES) and children's sleep in association with parent-reported externalizing and internalizing symptoms.

families ($p = .01$; Figure 3B). At lower levels of sleep problems, predicted means for internalizing symptoms were almost identical for children from higher and lower SES families. Conversely, at higher levels of sleep problems, children from lower SES families had higher predicted means for internalizing symptoms than their higher SES counterparts. We found no evidence for the prediction of internalizing symptoms by objective measures or their interactions.

Child-reported internalizing symptoms. None of the demographic characteristics were significant predictors of children's self-reported internalizing symptoms and, taken together, accounted for 6% of variance. Parent-reported sleep was not a significant predictor of internalizing symptoms. Of the actigraphy measures, sleep schedule variability was positively related to internalizing symptoms ($\beta = .32$; $p < .001$) and accounted for 10% of variance beyond demographics. As children's sleep schedules become less regimented and predictable, they report more internalizing symptoms. No interactions were found to qualify this result.

Longitudinal Prediction of Child Adjustment

Externalizing symptoms. Prior externalizing symptoms ($\beta = .69$; $p < .001$), race ($\beta = -.20$; $p = .003$), and SES ($\beta = -.21$; $p = .003$) were related to T2 externalizing symptoms. Demographic characteristics and the autoregressive effect accounted for 52% of the variance. Children's subjective reports of sleep were not directly related to externalizing symptoms; however, one interaction was detected between children's sleep–wake problems and race ($\beta = .21$; $p = .03$). AA children's externalizing symptoms did not change significantly along with changes in sleep problems but as EA children experienced more sleep problems they decreased on externalizing problems ($p = .05$; Figure 4). Notably, when sleep–wake problems were low, both EA and AA children had similar amounts of externalizing symptoms. However, when sleep–wake problems were high, AA children had more symptoms. This finding accounted for 2% of unique variance.

In using objective measures of sleep to predict externalizing symptoms, only sleep efficiency was significant ($\beta = .18$; $p = .008$). Surprisingly, increases in sleep efficiency predicted increases in externalizing symptoms and accounted for 2% of variance beyond demographics. No interactions were found in predicting externalizing symptoms from objective measures.

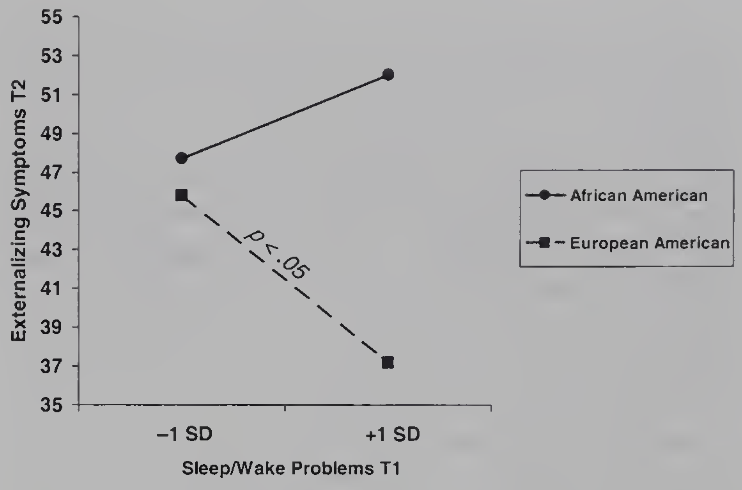


Figure 4. The interaction between race and sleep–wake problems at T1 in the prediction of parent-reported externalizing symptoms at T2.

Parent-reported internalizing symptoms. Prior internalizing symptoms ($\beta = .61$; $p < .001$), race ($\beta = -.20$; $p = .03$), and gender ($\beta = .15$; $p = .04$) were related to T2 internalizing symptoms. This meant that EA children and boys had higher parent-rated internalizing symptoms than AA children and girls, respectively. Demographic characteristics and the autoregressive effect accounted for 41% of the variance. Neither subjective report of sleep was related to internalizing symptoms. There was an interaction between sleep–wake problems and race ($\beta = .20$; $p = .05$), indicating that slopes for AA and EA children were significantly different from one another; however, neither slope was different from zero. This interaction accounted for only a very small amount of unique variance (1%), had no significant slopes, and resembles Figure 4 in terms of directions of slopes, and is not depicted. Importantly, and similar to the interaction evident for externalizing symptoms (Figure 4), predicted means for internalizing symptoms for AA and EA children were very similar at lower levels of sleep–wake problems. However, at higher levels of sleep problems, AA children appeared to have higher predicted means for internalizing symptoms in comparison to their EA counterparts.

None of the actigraphy sleep measures were directly related to internalizing symptoms. One interaction was found in which the link between sleep efficiency and internalizing symptoms was moderated by race ($\beta = .21$; $p = .02$). Sleep efficiency was unrelated to internalizing symptoms for EA children but was positively related to internalizing symptoms for AA children ($p = .01$). This interaction accounted for 2% of unique variance and is depicted in Figure 5.

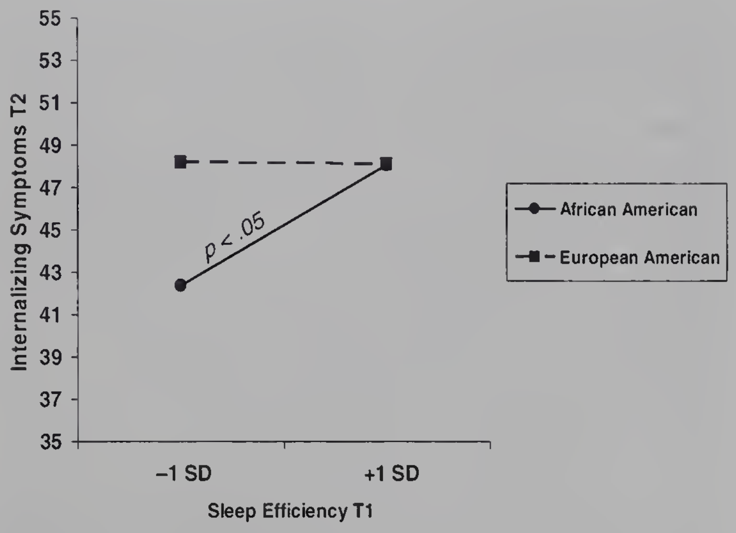


Figure 5. The interaction between race and sleep efficiency at T1 in the prediction of parent-reported internalizing symptoms at T2.

Discussion

A complex picture has emerged from predicting children's concurrent adjustment and changes in adjustment from both subjective reports of sleep from parents and children, objective actigraphy measures of sleep, and interactions between sleep and either race or SES. Some direct associations were observed between sleep characteristics and child adjustment. Further, different patterns of relations (i.e., moderation effects) were seen for children of different race and SES backgrounds. Interactions testing the study hypotheses accounted for relatively small but significant amounts of variance in children's internalizing and externalizing symptoms, ranging between 1% and 10%. These findings build on the literature and indicate that disruptions in children's sleep have a negative impact on multiple developmental processes over time and may depend on race and ecological context (e.g., Buckhalt et al., 2009; Gregory et al., 2005). In addition to establishing cross-sectional and longitudinal links between sleep and adjustment, the findings are unique in demonstrating that race and SES functioned as moderators of relations between children's sleep and adjustment. Results suggest that health disparities (Carter-Pokras & Baquet, 2002) need to be taken into account when examining child adaptation, and illustrate that certain populations may be more negatively impacted by adverse conditions (Foundation for Child Development, 2008). In this study, sleep was associated differentially with child adjustment based on the child's cultural or economic milieu. We found no evidence, however, for our exploratory hypothesis that gender may also moderate associations between sleep and adjustment. Importantly, in investigating the role of race and SES as moderators, any predictive ability of race also controlled for SES and vice versa.

Children's sleep was stable over 2 years: *r*s = .47 for sleep duration; .41 for sleep efficiency; .58 for parent-reported sleep problems; .40 for self-reported sleep–wake problems; and .19 for self-reported sleepiness. A few studies have documented that sleep problems persist across childhood (Jenni, Fuhrer, Iglowstein, Molinari, & Largo, 2005). In a 12-year longitudinal study of subjective measures of sleep, Gregory and O'Connor (2002) reported that sleep problems at age 4 persisted through childhood into midadolescence. Our findings parallel earlier findings and further highlight the stability of objective and subjective sleep parameters over 2 years of development, which has important implications for prevention and intervention considering associations between sleep and adjustment difficulties.

In original cross-sectional assessments when children were in third grade (El-Sheikh, Buckhalt, Cummings, et al., 2007), direct associations between various sleep parameters and child behavioral and emotional problems were found. The present results show that associations are still present 2 years later. Strengthening these findings is the fact that child-reported sleep problems were associated with parent-reported outcomes, indicating that these results are not due to mono-informant effects.

Importantly, the longitudinal results are novel and demonstrate the important role of sleep at T1 in predicting *changes* in child adjustment outcomes over 2 years, while controlling for T1 outcomes. The use of multiple reporters of children's adjustment, stringent control criteria, and assessments of objective and subjective sleep parameters increase confidence in our findings. In another report with the present sample (Buckhalt et al., 2009), sleep was a robust predictor of changes in children's cognitive and academic functioning examined through a comprehensive and standardized test battery (Woodcock–Johnson III Tests of Cognitive Abilities; Woodcock, McGrew, & Mather, 2001) and standardized achievement tests (Stanford Achievement Test [SAT], 2005). The pattern of effects in these studies illustrate that the negative effects of sleep disruptions are more pronounced for either AAs or children from lower SES homes across multiple domains of child development. A similar finding has been reported recently in which the link between adolescents' sleepiness and parent-reported executive functioning was more pronounced in the context of lower caregiver education (Anderson, Storfer-Isser, Taylor, Rosen, & Redline, 2009).

Race and SES moderated several associations between children's sleep and adjustment cross-sectionally. In relation to adjustment outcomes, AA and lower SES children tended to be more negatively influenced by sleep disruptions in comparison with EA and higher SES children. The most consistent pattern of moderation occurred cross-sectionally with SES. Children at lower levels of SES were well above average for externalizing symptoms, especially when they were also lower on sleep efficiency. As sleep efficiency increased, however, lower SES children dropped into more normative levels of externalizing symptoms. A similar result was seen with SES and child-reported sleep–wake problems; at higher levels of sleep–wake problems, parents of lower SES children reported more child internalizing symptoms. These moderated influences may not have long-lasting consequences, as we failed to find them when predicting future adjustment, although it should be noted that in this study the cross-sectional aspect of our study was conducted at T2 rather than T1; nevertheless, cross-sectional findings from T1 are supportive of this general pattern of SES-moderation effects.

When predicting future adjustment, the picture is slightly more mixed. It was evident that when child-reported sleep–wake problems were high, it was AA children who were reported by their parents as having the highest levels of externalizing and internalizing symptoms. In contrast, however, it was when objectively measured sleep efficiency was lowest that parents of AA children reported the lowest levels of internalizing symptoms. This finding is important for considering discrepancies between subjective and objective measures of sleep. Discrepancies between these categories of measures have been frequently noted by sleep researchers, and the general consensus is that we should continue to explore their different associations with outcome measures (Sadeh, 2008). Overall, effect sizes for moderation analyses were small to moderate, and interpretations of moderation showed a pattern of consistency cross-sectionally for SES and across time for race.

These findings suggest that negative outcomes associated with disruptions in a fundamental biological regulation system are exacerbated in the context of economic adversity or racial and ethnic minority status. The implication is that prevention and intervention efforts for children need to be tailored based on the family's resources and the challenges facing them. Periodic or chronic problems in children's sleep may have different associations with multiple outcomes depending on the family's circumstances. More generally, any single new problem facing a minority or low-SES family, such as illness of any member or a new financial challenge (to name but a few), may be enough to act as a "tipping point" toward broad dysfunction that may have various manifestations, including academic or behavioral problems in children. Alternatively, for families with more material and social resources, a single new challenge may have a lower level of impact. It is in this sense that the present findings are consistent with what others have termed the health disparities view (Carter-Pokras & Baquet, 2002).

Findings need to be interpreted within the study's particular context and limitations. Although sleep was examined objectively, via actigraphy, and subjectively, via both parent and child reports, these measures have limitations (Sadeh, 2008) including the inability to detect various sleep stages. Another study feature was the assessment of children's sleep and adjustment when they were in third and fifth grades. Assessments over other developmental periods may yield a different pattern of effects, and it is imperative for future research to examine developmental trajectories of child adaptation and sleep at various ages. Whereas we did not find gender-related moderation effects, such effects may be found in older children and adolescents. Although the present longitudinal study with two assessments is an advance in this new discipline of inquiry, full explication of developmental trajectories requires up to four assessments for the analysis of linear and nonlinear effects over time. Further, the causal relations between sleep problems and children's well-being appear to be bidirectional (e.g., Sadeh & Gruber, 1998). Just as sleep could be a symptom of child psychopathology or a significant stressor in a child's life, psychopathology could result from or be exacerbated by insufficient sleep and consequent fatigue and sleepiness, and establishing directionality of effects is a very important empirical question. It is also possible that the directionality of effects is different for children in various developmental periods, and thus models that test these various possibilities over a larger age span, and investigate alternative explanations, are critical. Furthermore, the current sample only included AA and EA children from the rural Southeast United States, and it is possible that regional customs with regard to sleep influenced the results. In addition, although we examined race and SES as moderators, there are still many questions to answer regarding the underlying processes by which they influence children's adjustment.

In summary, the current study's primary contribution is in demonstrating cross-sectional and longitudinal relations between various sleep parameters and externalizing and internalizing symptoms in a community sample of children. Findings highlight the importance of sleep quality and consistency of sleep schedule for children's optimal adjustment. The moderation effects of race and SES are intriguing and deserve continued attention. Interventions to improve sleep may yield significant benefits, and although benefits may accrue to all children, they may be especially valuable for children who are at greater risk for maladjustment and underachievement.

References

Acebo, C., Sadeh, A., Seifer, R., Tzischinsky, O., Wolfson, A., Hafer, A., et al. (1999). Estimating sleep patterns with actigraphy monitoring in children and adolescents: How many nights are necessary for reliable measures? *Sleep*, *22*, 95–103.

Adam, E. K., Snell, E. K., & Pendry, P. (2007). Sleep timing and quantity in ecological and family context: A nationally representative time-diary study. *Journal of Family Psychology*, *21*, 4–19.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Anderson, B., Storfer-Isser, A., Taylor, H. G., Rosen, C. L., & Redline, S. (2009). Associations of executive function with sleepiness and sleep duration in adolescents. *Pediatrics*, *123*, e701–e707.

Arbuckle, J. L. (2007). *Amos 16.0 user's guide*. Spring House, PA: Amos Development Corporation.

Aronen, E. T., Paavonen, E. J., Fjallberg, M., Soininen, M., & Torronen, J. (2000). Sleep and psychiatric symptoms in school-aged children. *Journal of the American Academy of Child and Adolescent Psychiatry*, *39*, 502–508.

Buckhalt, J. A., El-Sheikh, M., & Keller, P. (2007). Children's sleep and cognitive functioning: Race and socioeconomic status as moderators of effects. *Child Development*, *78*, 213–231.

Buckhalt, J. A., El-Sheikh, M., Keller, P. S., & Kelly, R. J. (2009). Concurrent and longitudinal relations between children's sleep and cognitive functioning: The moderating role of parent education. *Child Development*, *80*, 875–892.

Carter-Pokras, O., & Baquet, C. (2002). What is a "health disparity"? *Public Health Reports*, *17*, 426–434.

Chervin, R. D., Dillon, J. E., Archbold, K. H., & Ruzicka, D. L. (2003). Conduct problems and symptoms of sleep disorders in children. *Journal of the American Academy of Child & Adolescent Psychiatry*, *42*, 201–208.

Chorney, D. B., Detweiler, M. F., Morris, T. L., & Kuhn, B. R. (2008). The interplay of sleep disturbance, anxiety, and depression in children. *Journal of Pediatric Psychology*, *33*, 339–348.

Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology*, *112*, 558–577.

Crosby, B., LeBourgeois, M. K., & Harsh, J. (2005). Racial differences in reported napping and nocturnal sleep in 2-to-8-year-old children. *Pediatrics*, *115*, 225–232.

Dahl, R. E. (1996). The regulation of sleep and arousal: Development and psychopathology. *Development and Psychopathology*, *8*, 3–27.

Durrence, H. H., & Lichstein, K. L. (2006). The sleep of African-Americans: A comparative review. *Behavioral Sleep Medicine*, *4*, 29–44.

El-Sheikh, M., & Buckhalt, J. A. (2005). Vagal regulation and emotional intensity predict children's sleep problems. *Developmental Psychophysiology*, *46*, 307–317.

El-Sheikh, M., Buckhalt, J. A., Cummings, E. M., & Keller, P. S. (2007). Sleep disruptions and emotional insecurity are pathways of risk for children. *Journal of Child Psychology and Psychiatry*, *48*, 88–96.

El-Sheikh, M., Buckhalt, J. A., Keller, P. S., Cummings, E. M., & Acebo, C. (2007). Child emotional insecurity and academic achievement: The role of sleep disruptions. *Journal of Family Psychology*, *21*, 29–38.

El-Sheikh, M., Buckhalt, J., Mize, J., & Acebo, C. (2006). Marital conflict and disruption of children's sleep. *Child Development*, *77*, 31–43.

Entwisle, D. R., & Astone, N. M. (1994). Some practical guidelines for measuring youth's race/ethnicity and socioeconomic status. *Child Development*, *65*, 1521–1540.

Foundation for Child Development. (2008). Racial-ethnic inequality in child well-being from 1985-2004: Gaps narrowing, but persist. In D. J. Hernandez, & S. E. Macartney, *FCD Policy Brief, Child well-being index (CWI), #9*. Retrieved March 7, 2010, from http://www.ncmhd.nih.gov/child%20health.pdf

Fredricksen, K., Rhodes, J., Reddy, R., & Way, N. (2004). Sleepless in Chicago: Tracking the effects of adolescent sleep loss during the middle school years. *Child Development*, *75*, 84–95.

Gau, S. F., & Soong, W. T. (1995). Sleep problems in junior high school students in Taipei. *Sleep*, *18*, 667–673.

Gregory, A. M., Caspi, A., Eley, T. C., Moffitt, T. E., O'Conner, T. G., & Pouton, R. (2005). Prospective longitudinal associations between persistent sleep problems in childhood and anxiety and depression disorders in adulthood. *Journal of Abnormal Child Psychology, 33*, 157–163.

Gregory, A. M., & Eley, T. C. (2005). Sleep problems, anxiety and cognitive style in school-aged children. *Infant and Child Development, 14*, 435–444.

Gregory, A. M., & O'Connor, T. G. (2002). Sleep problems in childhood: A longitudinal study of developmental change and association with behavioral problems. *Journal of the American Academy of Child & Adolescent Psychiatry, 41*, 964–971.

Harter, S. (1982). The Perceived Competence Scale for Children. *Child Development, 53*, 87–97.

Harter, S. (1985). *Manual for the Self-Perception Profile for Children*. Denver, CO: University of Denver.

Hollingshead, A. B. (1975). *Four factor index of social status*. Unpublished manuscript.

James, E. M., Reynolds, C. R., & Dunbar, J. (1994). Self-report instruments. In T. H. Ollendick, N. J. King, & W. Yule (Eds.), *International handbook of phobic and anxiety disorders in children and adolescents* (pp. 317–330). New York: Springer.

Jenni, O. G., Fuhrer, H. Z., Iglowstein, I., Molinari, L., & Largo, R. H. (2005). A longitudinal study of bed sharing and sleep problems among Swiss children in the first 10 years of life. *Pediatrics, 115*, 233–240.

Kovacs, M. (1985). The Children's Depression Inventory (CDI). *Psychopharmacology Bulletin, 21*, 995–998.

Lee, K. A., McEnany, G., & Weekes, D. (1999). Gender differences in sleep patterns for early adolescents. *Journal of Adolescent Health, 26*, 16–20.

Little, R. J., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York: Wiley.

Milan, S., Snow, S., & Belay, S. (2007). The context of preschool children's sleep: Racial/ethnic differences in sleep locations, routines, and concerns. *Journal of Family Psychology, 21*, 20–28.

Must, A., Dallal, G. E., & Dietz, W. H. (1991). Reference data for obesity: 85th and 95th percentiles of body mass index (wt/ht^2) and triceps skin-fold thickness. *American Journal of Clinical Nutrition, 53*, 839–846.

Muzur, A., Pace-Schott, E. F., & Hobson, J. A. (2002). The prefrontal cortex in sleep. *Trends in Cognitive Sciences, 6*, 475–481.

Owens, J. A., Spirito, A., & McGuinn, M. (2000). The Children's Sleep Habits Questionnaire (CSHQ): Psychometric properties of a survey instrument for school-aged children. *Sleep, 23*, 1–9.

Paavonen, E. J., Almqvist, F., Tamminen, T., Moilanen, J., Piha, J., Rasanen, E., et al. (2002). Poor sleep and psychiatric symptoms at school: An epidemiological study. *European Child and Adolescent Psychiatry, 11*, 10–17.

Petersen, A. C., Crockett, L., Richards, M., & Boxer, A. (1988). A self-report measure of pubertal status: Reliability, validity, and initial norms. *Journal of Youth and Adolescence, 17*, 117–133.

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics, 31*, 437–448.

Reynolds, C. R., & Richmond, B. O. (1978). What I think and feel: A revised measure of children's manifest anxiety. *Journal of Abnormal Child Psychology, 6*, 271–280.

Rosen, C. L., Larkin, E. K., Kirchner, H. L., Emancipator, J. L., Bivins, S. F., Surovec, S. A., et al. (2003). Prevalence and risk factors for sleep-disordered breathing in 8- to 11-year-old children: Association with race and prematurity. *Journal of Pediatrics, 142*, 383–389.

Sadeh, A. (2008). Commentary: Comparing actigraphy and parental report as measures of children's sleep. *Journal of Pediatric Psychology, 33*, 406–407.

Sadeh, A., & Gruber, R. (1998). Sleep disorders. In A. S. Bellack & M. Hersen (Eds.), *Comprehensive clinical psychology* (pp. 629–653). New York: Pergamon.

Sadeh, A., Gruber, R., & Raviv, A. (2002). Sleep, neurobehavioral functioning, and behavior problems in school-age children. *Child Development, 73*, 405–417.

Sadeh, A., Raviv, A., & Gruber, R. (2000). Sleep patterns and sleep disruptions in school-age children. *Developmental Psychology, 36*, 291–301.

Sadeh, A., Sharkey, K. M., & Carskadon, M. A. (1994). Activity-based sleep-wake identification: An empirical test of methodological issues. *Sleep, 17*, 201–207.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.

Snedecor, G. W., & Cochran, W. G. (1967). *Statistical methods* (6th ed.). Ames: Iowa State University Press.

Stamatakis, K. A., Kaplan, G. A., & Roberts, R. E. (2007). Short sleep duration across income, education, and race/ethnic groups: Population prevalence and growing disparities during 34 years of follow-up. *Annals of Epidemiology, 17*, 948–955.

Stanford Achievement Test (10th ed.). (2005). San Antonio, TX: Harcourt Assessment.

Wirt, R. D., Lachar, D., Klinedinst, J. K., & Seat, P. S. (1990). *Personality Inventory for Children–Revised*. Los Angeles: Western Psychological Services.

Wisniewski, J. J., Mulick, J. A., Genshaft, J. L., & Coury, D. L. (1987). Test-retest reliability of the Revised Children's Manifest Anxiety Scale. *Perceptual and Motor Skills, 65*, 67–70.

Wolfson, A. R., & Carskadon, M. A. (1998). Sleep schedules and daytime functioning in adolescents. *Child Development, 69*, 875–887.

Woodcock, R. W., McGrew, K. S., & Mather, M. (2001). *Woodcock-Johnson III Tests of Cognitive Ability*. New York: Riverside.

Child Development, May/June 2010, Volume 81, Number 3, Pages 884–897

Experience-Based and On-Line Categorization of Objects in Early Infancy

Marc H. Bornstein and Clay Mash
Eunice Kennedy Shriver *National Institute of Child Health and Human Development*

What processes do infants employ in categorizing? Infants might categorize on line as they encounter category-related entities; alternatively, infants might depend on prior experience with entities in formulating categories. These alternatives were tested in forty-four 5-month-olds. Infants who were familiarized in the laboratory with a category of never-before-seen objects subsequently treated novel objects of the same category as familiar—they categorized on line—just as did infants who were exposed to objects from the same category at home for 2 months leading to their laboratory assessment of object categorization. Infants with home experience also recognized novel category objects as familiar from the outset—that is, prior experience with category exemplars was brought to bear in laboratory tasks.

One of the central challenges of object identification and categorization is the ability to recognize and organize category-related objects independent of their orientation, where they appear in the visual field, and so forth. For object perception and cognition to proceed with any degree of organization, order, or coherence, both biological and environmental variation must be transcended. Categorization constitutes a critical and adaptive solution to the challenges posed by both biological instability and environmental diversity. Categories refer to shared representations of like, but (usually) discriminable, entities and are in evidence when organisms treat (discriminably) different entities equivalently. Categorization allows generalization across experience, because categorization relates experienced entities to one another and to their mental representation. Categories mediate our interactions with the world in many ways (Bornstein, 1984; Harnad, 1987; Rakison & Oakes, 2003; Rosch, 1975; Smith, 1989; Smith & Medin, 1981). They structure and clarify perception and cognition, facilitate the storage and retrieval of information, and supply a principle of organization by which new information is banked efficiently in memory. An entity that is encountered for the first time could be treated as unique or, if categorization is brought to bear, in the same way as previously encountered entities that it resembles. Categorization is also used to generalize about newly encountered entities from the same category (Furrer & Younger, 2008; Gelman, 1988; Kalish & Gelman, 1992; Lin & Murphy, 2001; Proffitt, Coley, & Medin, 2000). Knowledge about a category affords expectations about novel entities in toto as well as about features of novel entities (Baldwin, Markman, & Melartin, 1993; Greco, Hayne, & Rovee-Collier, 1990; Mandler, 1998, 2000; Needham, Cantlon, & Ormsbee Holley, 2006). Categorization is thus a fundamental process by which the brain assigns meaning to sensory stimuli, and categorization is critical for rapidly and appropriately selecting behavioral responses (Freedman & Assad, 2006).

Little wonder, then, that the primate visual system has evolved such that single neurons respond selectively and quickly to complex object categories, such as faces and hands (Bruce, Desimone, & Gross, 1981). The activity of even tiny populations of neurons (on the order of approximately 100 randomly selected cells) over very short time spans (as brief as 12.5 ms) process accurate and robust information about both an object's identity and its category, and that category information generalizes over a range of object orientations, locations, and scales, even for novel objects (Freedman, Riesenhuber, Poggio, & Miller, 2001; Hung, Kreiman, Poggio, & DiCarlo, 2005; Schwarzlose, Swisher, Dang, & Kanwisher, 2008). Notably, specific category-related brain activity has been

---

We thank M. Arterberry, N. Cummins, J. Pugatch, E. Seiver, E. Stine, M. Stevenson, and T. Taylor. This research was supported by the Intramural Research Program of the NIH, NICHD.

Correspondence concerning this article should be addressed to Marc H. Bornstein, Child and Family Research, Program in Developmental Neuroscience, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, Suite 8030, 6705 Rockledge Drive, Bethesda, MD 20892-7971. Electronic mail may be sent to Marc_H_Bornstein@nih.gov.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8103-0016

observed in young human infants (Quinn, Westerlund, & Nelson, 2006).

Given the significance of categorization, surprisingly little is understood about the conditions required or the time course over which categorical representations are first formed in infancy. After all, infants are limited in the universe of objects to which they are exposed as well as in their cognitive capacities. Examining categorization processes in infancy can inform about factors that influence the nature and origins of categorization and the earliest accumulation of category-relevant knowledge (Furrer & Younger, 2008; Rakison & Oakes, 2003).

In a typical categorization study, infants are familiarized with discriminable exemplars of a category, and they are then tested with novel exemplars from the familiar category and from a contrasting different category (e.g., Arterberry & Bornstein, 2002; Behl-Chadha, 1996; Bornstein, Kessen, & Weiskopf, 1976; Quinn, Eimas, & Rosenkrantz, 1993). To show that they categorize, infants must include discriminably different novel exemplars of the category in the representation they have and simultaneously exclude contrasting novel exemplars from outside the category. Operationally, generalization of familiarization to a novel exemplar from the familiar category, coupled with heightened responsiveness to a novel exemplar from the novel category, is taken as evidence of category formation. For example, 5-month-olds who were habituated to varying facial expressions from one category (smiles) responded to a novel test image from the habituation category as if it were familiar, but responded differently to a novel (fearful) expression from a different category (Bornstein & Arterberry, 2003).

Routinely in infancy studies of object categorization, images or replicas of real-world objects are used, such as animals or vehicles or furniture, which are known and with which infants normally have experience. Visual-fixation studies have shown that infants as young as 3–4 months distinguish between categories such as animals and furniture, and even subcategories within animals and furniture such as cats, dogs, fish, and horses (Eimas & Quinn, 1994; Eimas, Quinn, & Cowan, 1994) or chairs, beds, sofas, and tables (Behl-Chadha, 1996). In these common cases, infants may categorize on the basis of their experience seeing exemplars in the course of the study or on the basis of experience they have in real life or from picture books, videos, television, and the like. Thus, it is unclear whether infants' categorization performance is based on preexisting familiarity with objects or categories, or, instead, if infants form categories during the (usually highly structured) experimental familiarization period. This circumstance leaves open key questions of whether and how infants actually construct categorical representations of objects they experience during fleeting encounters in routine daily life. Answers to these questions are important to fully understand human learning and how it develops.

Are object category representations in infants constructed before an experiment, or are they formed during the course of an experiment? If infants form category representations before being brought to the laboratory, it is probable that infants recognize experimental exemplars as representations of objects in the world with which they are already familiar and for which they may have some existing category knowledge. On this view, the familiarization phase of the familiarization/novelty-test procedure presumably serves to prime the knowledge that infants have already acquired outside the laboratory. If infants initiate category construction during the course of familiarization, then the familiarization/novelty-test procedure is better understood in terms of category formation. Infants presumably construct a category representation as they see exemplars from the familiar category. However, from the extant literature, it is difficult to completely rule out the possibility that prior experience or knowledge does not influence infant categorization. For most experiments designed to study categorization, it is possible that young infants recruit from their preexisting knowledge base in a way that, at least in part, determines their subsequent categorization behavior (Quinn, 2004).

In the present study, categorization was called *on line* when infants appeared to rely only on their immediate perceptions of the given experimental entities to form an object category within the familiarization period. Categorization was called *experience based* when infants could refer to some already existing experience, knowledge, or representation that was activated by perceptions of given entities but included properties other than those corresponding to the physical attributes of the experienced material. This study investigated whether infants younger than 6 months of age show on-line or experience-based categorization using an object-examination paradigm and objects that infants could never have seen before; therefore, the study controlled experience and eliminated preexisting knowledge from explaining infant object categorization performance.

Infants might show on-line categorization, or infants might show experienced-based categorization.

To take these two main options in reverse order, the second possibility holds that categorization in infants is experience based if infants refer to already existing knowledge representations that are being reactivated. By the middle to end of the 1st year (to take a common age span at which object categorization studies are performed), infants are already experience rich and perceptually competent, and so could easily be expected to take full advantage of opportunities to categorically structure their interactions with objects. There are three ways that preexisting experience could then affect categorization in the laboratory. One is that previous experience with the category is used during the familiarization phase of a categorization task and influences the acquisition or nature of the category representation. Another is that experience might be recruited during categorization judgments after the familiarization phase has ended. Familiarization might serve as a reminder that countervails forgetting by priming or reactivating an available but otherwise inaccessible category in memory (Tulving & Thomson, 1973), thereby facilitating access to that memory at the time of the subsequent test (Spear, 1973, 1978). In infants, reminders are effective in counteracting forgetting after retention intervals as long as 1 month (Fagen & Rovee-Collier, 1983; Rovee-Collier, Sullivan, Enright, Lucas, & Fagen, 1980; Sullivan, 1982). Last, of course, infants could use existing knowledge in both ways.

For example, asymmetries in young infants' categorization have been attributed to infants' relatively greater experience with one or another category (Quinn, 2002). Preliminary evidence for the role of previous experience in categorization was provided by Quinn, Yahr, Kuhn, Slater, and Pascalis (2002). Using a visual familiarization task, they found that 3- to 4-month-old infants with female primary caregivers formed an exclusive category of relatively unfamiliar male faces (i.e., a category that excluded female faces) but remembered individual exemplars from the more familiar category of female faces. Using an object-examining task, Pauen (2002) observed that 10- to 11-month-old infants responded to the distinction between furniture and animals both when between-category exemplars shared several perceptual properties and when they shared few. Pauen reasoned that on-line (perceptual) processes could not account for these results, and therefore infants' responding must reflect accumulated knowledge. Likewise, it has been suggested that categorization of cats and dogs differs for infants who do or do not have pets in their home. Kovack-Lesh, Horst, and Oakes (2005) reported that infants whose parents rated them as being highly interested in dogs performed best on a pattern-learning task involving pets.

Thus, experience and experience-dependent preferences seem to play a role in categorization performance (Needham et al., 2006, discuss related findings), and preexisting representations may help to explain categorization in young infants. More generally, these findings suggest that infants' experiences outside the laboratory can affect their categorical responding in the laboratory (Bar-Haim, Ziv, Lamy, & Hodes, 2006; Kelly et al., 2005, 2007).

However, other data cast some doubt on exclusively experienced-based interpretations of infant categorization. If infant categorization of nonhuman animals in the laboratory is assisted by their real-world experiences with nonhuman animals occurring before arrival at the laboratory, Quinn (2004) argued, one might expect having home pets to facilitate infants' animal categorization performance. Two studies compared animal categorization by infants who had been exposed to pets at home and infants who had not, and neither revealed a facilitative effect of home pets on categorization performance. Data from 3- to 4-month-old infants tested in Quinn's Experiment 1, and 6- to 7-month-old infants tested in his Experiment 2, offered no evidence that categorization of cats or dogs as measured in the laboratory was influenced by the presence of a home pet. Similarly, whether or not infants had a pet dog at home was not a significant predictor of their category task performance, nor did it correlate with infants' reported interest in dogs in Saffran, Pollak, Seibel, and Shkolnik (2007). "These null results thus fail to support the suggestion that infant categorization of nonhuman animals in the laboratory is assisted by real-world experience with nonhuman animals occurring prior to arrival at the laboratory" (Quinn, 2005, p. 113).

On the first possibility, then, infant categorization might not (solely) reflect their prior accumulated experience. Categorization might take place on line if infants rely on their immediate experiences with given stimuli to form categories. Even if infants possess some categorical knowledge, some infant categorization performance might still be best explained by spontaneous category learning (Mareschal, French, & Quinn, 2000; Mareschal & Quinn, 2001). For example, infants presented with the contrast between replica animals and people did not categorize in the first half of the task but gave evidence that they did in the second half (Oakes, Plumert, Lansink, & Merryman, 1996).

Infants in this study seemed to discover a categorical distinction as they actively explored the experimental objects. Younger and Fearing (2000) provided evidence for infant on-line category formation by showing that infants form categories at multiple levels that depend on similarities among the entities presented during the experiment. Presented with cat versus dog contrasts manipulated by changing featural (although not categorical) characteristics of the stimuli, infant responsiveness varied, suggesting that they formed categories during familiarization (and that category boundaries could be manipulated depending on the information presented during familiarization; see also French, Mermillod, Quinn, & Mareschal, 2001). The authors reasoned that if infants were tapping into category representations that were fixed and established before their participation in this task, infant responsiveness would not differ with featural variation in the category information presented. Together, these findings suggest that categorization performance in infancy can be attributed to active on-line processing and within-task learning.

Finally, infant performance need not be one (on line) or the other (experience based), but both. Four-month-olds' previous experience with dogs, cats, or both and their on-line looking behavior during learning an adult-defined category of cat in a visual familiarization task jointly determined whether or not infants categorized (Kovack-Lesh, Horst, & Oakes, 2008). Infants were divided by whether they had a pet at home and how much they shifted their gaze back and forth between stimuli during familiarization. Infants in a pet/high-switching group showed significantly greater novelty preference than did infants in a no-pet/high-switching group and infants in a pet/low-switching group. In this study, infants who had a pet at home and who engaged in relatively high levels of stimulus comparison demonstrated the highest level category functionality—they remembered individual items and they recognized commonalities among items from within the same category. Infants in other groups did not respond either to the category or to the individual items. One problem with this pattern of results, however, was that infants in the most advanced pet/high-switching group did not differ in novelty preference from infants in the least advanced no-pet/low-switching group.

The common use of real or familiar stimuli in these infancy studies restricts the analysis of performance contrasts and the important deductions that can be made from them. Even if young infants have not observed the exact stimuli in categorization experiments prior to their participation, they have been exposed to related stimuli that may be sufficient to build up representations that could influence learning within those tasks (Needham & Baillargeon, 1998, and Quinn & Eimas, 1998, discuss relations between real-world experience and within-task learning). It could also be that parents with pets treat their infants differently than do parents who do not have pets. In either case, infants might differ in their attention and learning in general. Investigating infants' categorization abilities, some researchers have used novel stimulus materials such as meaningless dot patterns and unusual creatures (e.g., Bomba & Siqueland, 1983; Younger & Cohen, 1983). However, dot patterns are simplified stimuli and hardly ecologically representative of 3-D objects that have affordances for action and serve functional roles, and hybrid animal-part studies depict familiar parts and/or part–whole relations that enable (or specifically draw on) linkages to previous experience. Many factors influence the ease of representation of the input and facilitate categorization, and one obvious factor is familiarity; elements similar to those that have been encountered previously can be expected to be easier to represent, maintain in memory, and perhaps categorize (Saffran et al., 2007). Thus, materials that infants can readily represent, such as animals, vehicles, and furniture, might facilitate categorical task performance. Finally, one can expect natural variation in infants' experiences and, in consequence, infant performance in any natural experiment through unsystematic exposure to materials and images.

How then do infants categorize objects? To address this question, task performance is best examined independently of differing levels of stimulus experience and familiarity. Here, we investigated infant object categorization by directly examining infants' learning of entirely new object categories both with and without previous experience with object exemplars of those categories. Experience was controlled through the development and use of purpose-constructed never-before-seen objects and categories, and whether or not infants were shown the objects preceding their participation in a laboratory categorization task. If infants with no possible experience with such an object category and its exemplars form a categorical representation during a laboratory task, then previous, cumulative experience must not be requisite for object categorization, and infants can categorize "on the fly." If infants with no experience with the object category and its exemplars cannot form a categorical representation during a laboratory task, then previous, cumulative experience may be requisite for object categorization. If infants having such previous, cumulative experience with the object category and its exemplars form a categorical representation during a laboratory task, then that experience would appear to define a milestone of experience-dependent sufficiency for learning new categories and subsequently categorizing novel exemplars. Thus, along with its control of experience with the object stimuli, this study offers an experimental vantage point from which to examine infants' developing categorization abilities. In this experiment, we contrasted laboratory object categorization in young preverbal infants given prior home experience with unique and entirely novel objects to laboratory object categorization in infants given no such experience.

Method

Infants

Nineteen normal and term infants (5 girls) were 3.15 months (*SD* = 0.26) when they began the home experience condition; they were 5.38 months (*SD* = 0.39) at the laboratory assessment. (One additional infant was recruited, but the family moved during the home experience phase.) Twenty-five normal and term 5.12-month (*SD* = 0.42) infants (13 girls) participated in the no home experience condition. (An additional 3 infants were recruited, but their data were not included due to fussiness or experimenter error.) There were no differences in sociodemographic characteristics of infants or families in the two conditions.

Stimuli

Never-before-seen objects were constructed to comprise two novel categories (Figures 1A and B) for infants. The objects were each made of three wooden pieces assembled in unique configurations. Individual parts were painted with a total of eight different colors, four used for each category. Each category exemplar represented a unique combination of part configuration and color. Both within and between categories, the parts of each object were assembled in a similar way, but each varied with respect to the exact positioning of parts. Each category contained a total of eight unique

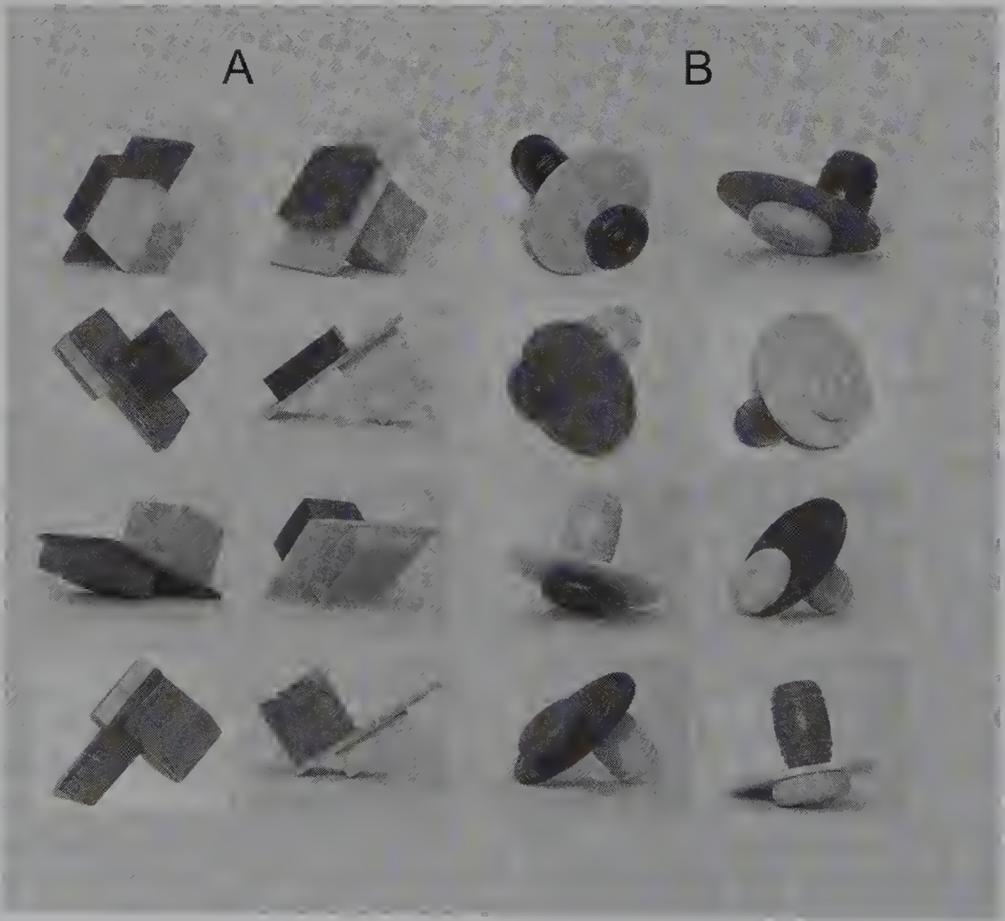


Figure 1. Examples of Categories A and B.

exemplars. Each object was painted with three colors (one per part), and the colors did not overlap across categories. Thus, the two categories were similar in size, complexity, part configuration, and the distribution of color boundaries but differed in the parts and colors of the objects. Curved versus linear edge features were distributed over the entire object for each exemplar and category. The familiarization exemplars were depicted in numerous and varied orientations and viewpoints, so edge composition did not provide a single, location-invariant feature for discrimination between images.

Home Experience Procedure

At the 3-month laboratory visit, parents of infants in the home experience condition received materials and training for conducting stimulus exposure sessions at home. They were provided with a booklet containing 20 photographs, four pictures each (differing in orientation) of five randomly selected exemplars from one category. The images were 10.2 cm × 15.2 cm and printed on standard size plain white paper, which was then laminated. Parents were instructed to show their infants the pictures once per day on a daily basis until they returned to the laboratory for the experimental task. They were provided with a timer and were instructed to pace presentations so that the entire stimulus set was shown over the course of 2 min. Parents were telephoned on a weekly basis to evaluate compliance with the procedures. Mothers were accommodating in conducting the stimulus exposure sessions at home, reporting successful administration of 5.54 sessions/week on average (*SD* = 1.04) throughout the 2-month home-experience period. They reported no significant problems with administration. Missed sessions were distributed uniformly over participants, most typically resulting from competing household demands and minor illness.

Laboratory Object Examination

Object examination procedure. To test preference, discrimination, and categorization, an object-examination procedure was used following established conventions (Oakes & Tellinghuisen, 1994; Ruff, 1986). Actual objects were presented directly to infants. The data of primary interest from object examination are the amounts of time infants spend in focused examination of objects while handling them. Oakes and Tellinghuisen (1994) have argued that focused examination provides a more sensitive index of information processing than traditional measures of looking time. Each session was videorecorded, and infants' examination of stimulus objects was coded from the videorecords. All coders were blind to the hypotheses of the study and condition assignment of the infants. Reliability of coding was established by calculating agreement between coders. All together, 35% of sessions were coded twice, with high agreement, $r = .91$.

Stimulus preference and discrimination. To assess baseline stimulus preference and within-category exemplar discrimination, an independent sample of 14 normal and term infants participated in a fixed-trials familiarization/novelty-preference task (*M* age = 5.23 months, *SD* = 0.26; 6 girls). (Two additional infants were tested, but their data were not included due to fussiness and inattention.) Infants participated in two different sessions, one involving each of the two different object categories. For each session, a single exemplar of one category was randomly selected and presented repeatedly over the course five 20-s familiarization trials. At the end of each trial, the experimenter retrieved the object from the infant, removed it from view, then re-presented it to begin the next trial. Two discrimination trials were then administered in the same fashion. In one trial the same object was presented again, and in the other trial a randomly selected novel member of the same category was presented. The order in which the discrimination trials were administered (novel exemplar first or second) was counterbalanced across infants. The order of sessions for the two categories was also counterbalanced.

To assess infants' baseline preference, infants' object examination was compared between categories within the first 20 s of exposure during the familiarization phase. Examination times did not differ between categories (A: $M = 2.88$ s, $SD = 4.37$ vs. B: $M = 3.66$ s, $SD = 4.80$), $t(13) = 0.70$, *ns*. Infants' discrimination of exemplars within categories was examined in an analysis of variance (ANOVA) utilizing category (A vs. B) and discrimination trial (familiar exemplar: $M = 2.05$ s, $SD = 1.65$, vs. novel exemplar: $M = 3.50$ s, $SD = 2.01$) as within-infants factors. Only the discrimination trial contrast was significant, $F(1, 13) = 6.10$, $p = .028$, with infants examining the novel exemplars longer than the familiar ones. In brief, infants had no a priori preference for one category over the other and could discriminate among same-category exemplars.

Categorization. All infants participated in the same laboratory categorization task. The categorization procedure comprised a familiarization phase and a test phase. Five objects from one (or the other) category were randomly assigned for familiarization. Using a fixed-trials procedure (Ruff, 1986), each object was examined twice over the course of 10 familiarization trials in an order uniquely randomized for each infant. During familiarization, the experimenter presented one object at a time to the infant, using a count-down timer to mark an elapsed time of 20 s per trial. At the end of each trial, the experimenter retrieved the object from the infant and then presented the next object. Except for the one being presented, the remaining objects were hidden from the infant's view. Two test trials were then administered in the same fashion. In one test trial a novel previously unseen member of the familiar category was presented, and in the other test trial a randomly selected member of the novel category was presented. The order in which the test trials were administered (novel category first or second) was counterbalanced over infants.

In summary, infants in the home experience condition and infants in the no home experience condition came to the laboratory at 5 months. At that laboratory visit, infants were familiarized to five objects—for infants in the home experience condition, they were the same objects as those depicted in the picture booklet the infants saw at home—and then tested with two previously unseen objects (one from each category). Half of each sample was familiarized to each category, after which every infant received both a familiar-category novel test trial and a novel-category novel test trial. Laboratory visits and data analysis were conducted in exactly the same manner for infants in the home experience and no home experience conditions.

Results

Infants' laboratory object-examination categorization was analyzed in two mixed-model ANOVAs utilizing familiar category (A vs. B) as a between-infants factor and a series of planned trial contrasts as within-infant factors. Planned comparisons examined effects of (a) familiarization (mean of first two familiarization trials vs. mean of last two familiarization trials), (b) category generalization (mean of last two familiarization trials vs. the familiar-category novel-object test trial), and (c) category discrimination (mean of last two familiarization trials vs. the novel-category novel-object test trial). Preliminary analyses revealed no effects of test condition order or gender, so subsequent analyses excluded those factors.

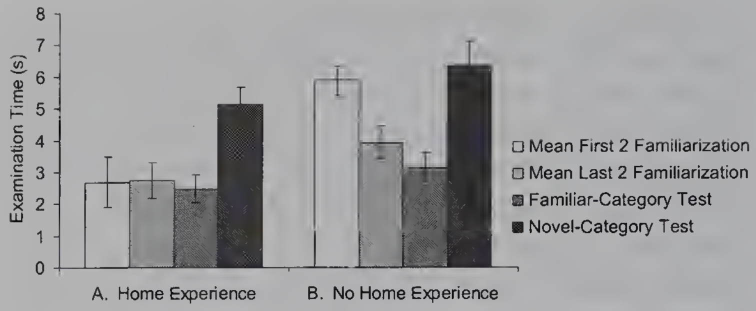


Figure 2. Mean examination time (in s ± 1 *SEM*) at the start of familiarization, at the end of familiarization, for the test of the familiar-category novel exemplar, and for the test of the novel-category novel exemplar: (A) home experience condition and (B) no home experience condition.

Familiarization, Category Generalization, and Category Discrimination Following Prior Home Experience

The familiarization and category generalization contrasts were not significant, *F*s(1, 18) = 0.009 and 0.334, *ns*, but the category discrimination contrast was significant, *F*(1, 18) = 14.05, *p* = .001, partial η^2 = .438 (Figure 2A).

Familiarization, Category Generalization, and Category Discrimination Without Prior Home Experience

The familiarization contrast was significant, *F*(1, 24) = 5.73, *p* = .025, partial η^2 = .193; the category generalization contrast was not significant, *F*(1, 24) = 1.21, *ns*; and the category discrimination contrast was significant, *F*(1, 24) = 7.92, *p* = .010, partial η^2 = .248 (Figure 2B).

Familiarization Differences and Similarities Between Experimental Conditions

To examine whether their home experience modified infants' familiarization (encoding) during the laboratory task, three additional analyses compared object examination times between the two groups. First, the means of the first two familiarization trials differed across conditions, *M* = 2.69 s (*SD* = 2.09) in the home experience condition versus *M* = 5.88 s (*SD* = 4.91) in the no home experience condition, *F*(1, 42) = 7.01, *p* = .011, partial η^2 = .143. Second, a main effect of condition emerged over the start through the end of familiarization, *F*(1, 42) = 5.02, *p* = .030, partial η^2 = .107, with infants in the no home experience condition examining objects longer ($M = 9.80$ s, $SD = 6.92$) than infants who had previous home experience ($M = 5.44$ s, $SD = 3.53$). Additionally, trial interacted marginally with condition, $F(1, 42) = 3.59$, $p = .065$, partial $\eta^2 = .079$. As noted earlier, a significant decrement was observed in the no home experience condition but not in the home experience condition. Third, the average duration of object examination per individual look throughout familiarization was significantly longer in the no home experience condition ($M = 2.48$ s, $SD = 1.96$) than in the home experience condition ($M = 1.41$ s, $SD = 1.07$), $F(1, 42) = 4.64$, $p = .037$, partial $\eta^2 = .099$.

To assess infants' general orientation to the task and its demands during familiarization, two additional analyses were conducted. First, the total duration of casual looking toward the stimuli—that is, general visual regard not limited to focused examination—was calculated across familiarization trials and compared between groups. Mean looking did not differ between infants in the home experience ($M = 12.10$ s, $SD = 2.58$) and no home experience ($M = 10.54$ s, $SD = 4.64$) conditions, $t(42) = 1.32$, *ns*. Additionally, general interest in the stimulus objects was independently rated for each trial using a 5-point Likert-type scale from 1 = *no interest* to 5 = *extremely interested*. Fifteen percent of the sessions were rated by a second coder, and ratings correlated $r = .94$ between coders. Ratings did not differ between infants in the home experience ($M = 3.75$, $SD = 0.50$) and no home experience ($M = 3.76$, $SD = 0.88$) conditions, $t(42) = 1.32$, *ns*. Together, these findings indicate that infants' general engagement in the task did not differ between conditions and therefore cannot account for the differences observed in infants' processing of the stimulus objects during familiarization.

Discussion

At 3 months, infants were randomly assigned either to a home experience condition that consisted of their parents showing them images of exemplars of a novel, never-before-seen category of objects or to a no home experience condition. At 5 months, all the same infants participated in a laboratory categorization study with objects from those same categories. Infants in the home experience condition did not decline in looking at objects over the course of laboratory familiarization, they generalized from the end of familiarization to presentation of a novel within-category object, and they examined a novel-category novel object longer than they examined objects at the end of familiarization. Infants in the no home experience condition declined in looking over the course of familiarization to the objects, they generalized from the end of familiarization to the presentation of a novel within-category object, and they examined a novel-category novel object longer than they examined objects at the end of familiarization. These findings were not ascribable to infants' differential engagement with the stimuli or to differential preferences for or failure to discriminate among the stimuli.

Two Lessons Learned

This study of infant categorization of ad hoc, unfamiliar objects eventuated in two main sets of findings. First, 5-month infants with no prior experience with a completely novel category of never-before-seen objects categorized them on line at the time of an experimental session, just as infants would presumably do in encountering category-related objects in everyday life. Second, infants with prior experience also categorized the objects at the time of an experimental session, and home experience with completely novel objects in a novel object category also influenced infant behavior at the time of the later experimental laboratory session.

Infants familiarized to exemplars of a never-before-seen object category during an experimental session generalized examination time to a novel object exemplar from the same category, while discriminating one from a contrasting category. Because otherwise equivalent infants showed no baseline preference for exemplars of one category over the other, and demonstrated an ability to discriminate exemplars within both categories, the no home experience infants' performance clearly reveals their ability to categorize objects on line, that is as never-before-seen objects in a category are encountered. Because infants participating in the no home experience condition had never previously experienced exemplars of the novel stimulus categories, their representations could only have formed during the task itself. These findings affirm the position that young preverbal infants' responses during laboratory tasks incorporate and reflect experience during the task. This study indicates that infants can formulate categorical representations of objects entirely within the confines of a few short minutes exposure. Infants appear to have learned an object category during familiarization. It may be that mere exposure on the task of comparing different exemplars of the same category over time leads to the extraction of a category

(Gentner & Medina, 1998). In adults, categorization experience with stimuli analogous to those used in testing attunes the individual and the brain to relevant dimensions of a stimulus (Baker, Behrmann, & Olson, 2002; Goldstone, Steyvers, & Larimer, 1996; Jiang et al., 2007; Schyns, Goldstone, & Thibaut, 1998).

This is not to imply that prior experience plays no role in infant cognition. The second question of whether infants construct categorical representations over their encounters with objects as they arise in everyday life and later call on that information was examined in the complementary home experience condition. These infants were exposed to pictures of novel objects every day (for just a few seconds each) over the course of 2 months. When these infants participated in the same laboratory categorization task, they too demonstrated categorical representations during the test phase of the task, but no evidence of learning during the familiarization phase. On the convention that attention during familiarization reflects active encoding (Bornstein, 1985; Colombo, 1993), experience at home clearly exerted a direct effect: More encoding occurred when prior opportunity for encoding was absent. This finding underscores the conclusion that infants utilized stimulus representations they had constructed at home, bringing them to bear on their categorization performance in the laboratory. Because the category exemplars were novel and unique to this research program, infants must have constructed categorical representations of the stimulus objects over the course of their distributed home experience, brought this information with them to the laboratory, and then used this information during the laboratory task. The fact that infants in the home experience condition (a) treated stimuli at the beginning and end of the laboratory familiarization the same and (b) treated stimuli at the start of familiarization differently from infants in the no home experience condition provides additional evidence that infants carried their knowledge of the category from previous home experience into the laboratory. (Checks verified that the two groups did not differ overall in engaging the stimuli.) This finding is consistent with the positions that infants can construct stable categorical representations of objects they experience and that those representations can guide the systematic generalization of familiarity and knowledge to novel entities. The different patterns of familiarization between groups for objects that were previously experienced versus those that were not indicate that prior representations (based on experience) exerted a more dramatic influence on property extensions than did in-task modeling.

Furthermore, young infants in the home experience condition had to transfer knowledge from pictures to objects, and from looking to looking and handling. These differences are nontrivial and require intersensory and mental generalization. This deduction accords with previous research that revealed the abilities of infants in this age range to detect equivalences between objects and their photographs (e.g., DeLoache, Pierroutsakos, & Uttal, 2003). Real-life exposure to images of objects appears to influence infants' responses to categories of the same three-dimensional objects in the laboratory.

Object Representation: Items or Prototypes?

On what basis do infants learn object categories? One approach to answering this question starts from the nature of infant representation in category development. Object categories may be represented by individual exemplars of the category or by a summary representation (perhaps a prototype or rule) of the category (e.g., Medin & Schaffer, 1978; Nosofsky, 1984). Item-based approaches tend to rely on (sometimes large) mnemonic stores of previous experience; prototype or rule-based approaches usually rely on simpler criteria and presumably do not require extensive memory. In adults, when the number of stimuli is small, categorization may be mediated by individual items, but when the number of stimuli exceeds some capacity, categorization is presumably handled by prototypes or rules rather than by individual exemplars (Ashby & Ell, 2001). It has been asserted that infants form categories both as collections of individual items (e.g., Hayne, Rovee-Collier, & Perris, 1987; Quinn et al., 2002; Younger, 1993) and as organized around prototypes (e.g., Bomba & Siqueland, 1983; Bornstein, 1982; Younger, 1985). Which infants do appear to depend on their age, the task at hand, or other factors. Younger (1990) reported that 13-month infants use specific-item information in categorization, but younger babies do not. Quinn (2002) proposed that experience with exemplars from a particular category changes infants' categorization from forming a summary representation (prototype) of exemplars to remembering each individual exemplar as well as maintaining a summary representation. Horst, Oakes, and Madole (2005) also reported results that support infants' learning individual exemplars and forming a summary representation. Perhaps after extensive experience with a limited number of items in a category, infants can both maintain a summary representation or prototype and remember individual items. In light of the constancy in infants' home experience in our study—seeing the same five objects each in a limited number of orientations every day—it is possible that infants stored and retained each exemplar individually and compiled a summary representation (prototype) of the category. If so, home experience infants' flat familiarization profile could reflect individual item recognition or common reference to a previously developed category (prototype) representation. It is impossible to discern from the present familiarization data which format infants used to encode their home experience and applied during laboratory familiarization.

That said, any related ambiguity with respect to infants' familiarization performance in no way undermines our interpretation of infants' test phase performance where both groups provided clear evidence for categorical representations. Because the home experience infants in this present study demonstrated categorization of familiar-category novel exemplars in the test, and there was no evidence for their learning during familiarization, we deduce that infants acquired a categorical representation prior to the laboratory task (i.e., over the course of home experience) and used it during the test. It is possible that infants in the home experience condition formulated specific memories of each individual image or stimulus object and showed reduced examining during familiarization because they recognized individual stimuli instead of recognizing the category. On closer inspection, however, this interpretation is not consistent with the evidence available. First, infants in the laboratory task had to bridge their home experience with pictures into their responses to objects over significant differences in setting, context, and presentation mode. More critically, however, if infants simply recognized individual objects during familiarization, they would not have generalized examination time to either test stimulus. Because infants in a separate task discriminated among members of both categories, the familiar-category novel-object test stimulus would have been discriminated if infants relied on item recognition alone. As reported earlier, however, this is not what happened. Infants in the home experience condition, like their no home experience counterparts, generalized looking from familiarization to the familiar-category novel object.

Object Representation: Features or Configurations?

Another approach to the basis of infant object categorization turns on the question of what about objects and object categories constitute critical criteria. Infants in our study treated discriminably different same-category novel objects in the test equivalently to the objects in the familiarization set and discriminated different-category novel objects. Some investigators contend that infants younger than 1 year of age form categories based on knowledge about causal and functional similarities of objects (Gelman, Durgin, & Kaufman, 1995; Mandler, 2000, 2004; Mandler & McDonough, 1993; McDonough & Mandler, 1998; Van de Walle, 1999). The object stimuli used in the present study lack causal or functional relations as such, and so these bases of categorization are ruled out. Thus, novel objects were presumably categorized or discriminated by their similarity of appearance to familiar ones. Such comparisons could refer to (a) similarity of critical features or (b) similarity of a configuration of features. In a sense, the answer to this question further specifies the nature of the infant's representation of the object category.

Perhaps infant object representations are holistic, based on an object category's overall configuration or appearance without selective attention to individual features, dimensions, or attributes. Using sequential touching, Rakison and Butterworth (1998) deduced that infants under 2 years attend, not to individual object parts per se, but to the structural configuration of the parts perceived as a whole, and Träuble and Pauen (2007) found that 1-year-olds categorized objects on the basis of their overall similarity and not their part similarity. Elsewhere, infants have been shown to rely on overall shape similarity as a basis of object categorization (Graham, Kilbreath, & Welder, 2004; Welder & Graham, 2001). Later, infant categorization could become more analytic, with selective attention paid to features, dimensions, or attributes (Shepp, 1978). There is evidence, too, however, that young infants discriminate and categorize on a so-called partonomic bias. Tversky and Hemenway (1984) contended that children first form categories based on partonomies rather than taxonomies (see also Mash, 2006; Rakison & Butterworth, 1998; Rakison & Cohen, 1999). This means that infant categorization refers to a focus on specific parts rather than the entire object (Booth & Waxman, 2002; Madole & Cohen, 1995; Mash, 2006). Parts are known to be highly relevant for object discrimination and categorization in adults as well as in children

(Biederman, 1993; Hummel & Biederman, 1992; Madole & Cohen, 1995; Saiki & Hummel, 1996; Tversky & Hemenway, 1984; Younger, 1990). Younger and Cohen (1983, 1986) found that 10-month-olds recovered attention on the basis of correlations among stimulus features, whereas 4- and 7-month-olds recovered attention on the basis of individual stimulus features. Children graduated over age from processing features (parts of animal pictures) independently at 4 months to processing correlations among features (the animal as a whole) at 10 months. Well into their 2nd year of life, infants seem still to reference prominent perceptual features of objects. Infants between 1 and 2 years categorize unmodified cars from unmodified cows (Rakison, 2000). However, 14-month-olds fail to categorize cows from cars when object parts (legs and wheels) are removed, and 14- and 18-month-olds fail to categorize cows from cars when the same object parts are possessed by all exemplars.

The stimuli in the present task held configural relations constant between categories (i.e., part attachment positions and angles), whereas the shape and color of local parts differed. Thus, for infants to categorize our stimuli, they had to compare the objects on the basis of part features (shape, color, or both). Because the familiarization and test objects in this study had similar perceptual features, infants' choice of object following familiarization likely reflects a perceptual match of the test object to the familiarization exemplars (Rakison, 2005; Rakison & Hahn, 2004; Rakison & Poulin-Dubois, 2001). It is also possible of course that local part differences give rise to overall shape differences if separate parts enter equally into visual analysis and memory. Thus, the perceptual basis of object categorization—features or configurations—appears still an open question.

Future Directions and Conclusions

Infants in a home experience condition saw a variety of objects for approximately 100 min in total (or 20 min per object image), which was apparently sufficient to influence their later laboratory familiarization behavior. Future research might be directed at quantifying exposure times more precisely and determining relations between amount and kind of exposure (e.g., massed vs. distributed practice) and the development and robustness of infants' resulting categorical representations. Questions about whether it is the amount, kind, or schedule of experience can be addressed in training studies similar to this one in which infants are assigned parametrically programmed protocols to determine the critical experiences for formation and use of object category knowledge. Extensive experience with items from a particular category likely help infants form relatively strong and stable category representations. Further comparisons could be made between visual only, manual only, and visual–manual experiences to determine whether intersensory information confers category construction advantages over single sensory experience alone (Iachini, Borghi, & Senese, 2008). Additional research might be directed to uncover how infants come to regard novel objects as familiar, what information they store about objects, and how object categories arise even earlier in life. Is experience with particular objects or with categories of objects critical? Metacognitively, an effect of experience need not be specific to a particular categorical contrast or set of stimuli; experience outside the laboratory with one set of categories might influence infants' responding to a new category in the laboratory.

The categories we used here were ad hoc and unfamiliar, and babies' reactions to them reveal a basic category abstraction mechanism in infancy that does not seem to require a deep understanding of any specific object domain. We note, too, that just because the stimuli are simple and abstract, it does not automatically follow that these results are generalizable. Nonetheless, an impressive category extraction mechanism seems to be at work in the first half-year of life that gives infants the wherewithal and flexibility to learn new categories on the fly as well as retain and apply those categories they have previously experienced. These findings articulate with, and bring to a functional behavioral level, evidence showing that distinct neural signals correspond with category learning and representation in infants as young as 6 months.

References

Arterberry, M. E., & Bornstein, M. H. (2002). Variability and its sources in infant categorization. *Infant Behavior and Development*, *25*, 515–528.

Ashby, F. G., & Ell, S. W. (2001). The neurobiology of human category learning. *Trends in Cognitive Sciences*, *5*, 204–210.

Baker, C. I., Behrmann, M., & Olson, C. R. (2002). Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nature Neuroscience*, *5*, 1210–1216.

Baldwin, D. A., Markman, E. M., & Melartin, R. L. (1993). Infants' ability to draw inferences about nonobvious object properties: Evidence from exploratory play. *Child Development*, *64*, 711–728.

Bar-Haim, Y., Ziv, T., Lamy, D., & Hodes, R. M. (2006). Nature and nurture in own-race face processing. *Psychological Science*, *17*, 159–163.

Behl-Chadha, G. (1996). Basic-level and superordinate-like categorical representations in early infancy. *Cognition*, *60*, 105–141.

Biederman, I. (1993). Visual object recognition. In A. I. Goldman (Ed.), *Readings in philosophy and cognitive science* (pp. 9–22). Cambridge, MA: Bradford Press.

Bomba, P. C., & Siqueland, E. R. (1983). The nature and structure of infant form categories. *Journal of Experimental Child Psychology*, *35*, 294–328.

Booth, A., & Waxman, S. (2002). Object names and object functions serve as cues to categories for infants. *Developmental Psychology*, *38*, 948–957.

Bornstein, M. H. (1982). Perceptual anisotropies in infancy: Ontogenetic origins and implications of inequalities in spatial vision. In H. W. Reese & L. P. Lipsitt (Eds.), *Advances in child development and behavior* (Vol. 16, pp. 77–123). New York: Academic Press.

Bornstein, M. H. (1984). A descriptive taxonomy of psychological categories used by infants. In C. Sophian (Ed.), *Origins of cognitive skills* (pp. 313–338). Hillsdale, NJ: Erlbaum.

Bornstein, M. H. (1985). Habituation of attention as a measure of visual information processing in human infants: Summary, systematization, and synthesis. In G. Gottlieb & N. A. Krasnegor (Eds.), *Measurement of audition and vision in the first year of postnatal life: A methodological overview* (pp. 253–300). Westport, CT: Ablex.

Bornstein, M. H., & Arterberry, M. E. (2003). Recognition, discrimination and categorization of smiling by 5-month-old infants. *Developmental Science*, *6*, 585–599.

Bornstein, M. H., Kessen, W., & Weiskopf, S. (1976). Color vision and hue categorization in young human infants. *Journal of Experimental Psychology: Human Perception and Performance*, *2*, 115–129.

Bruce, C., Desimone, R., & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *Journal of Neurophysiology*, *46*, 369–384.

Colombo, J. (1993). *Infant cognition: Predicting later intellectual functioning*. Newbury Park, CA: Sage.

DeLoache, J. S., Pierroutsakos, S. L., & Uttal, D. H. (2003). The origins of pictorial competence. *Current Directions in Psychological Science*, *12*, 114–118.

Eimas, P. D., & Quinn, P. C. (1994). Studies on the formation of perceptually based basic-level categories in young infants. *Child Development*, *65*, 903–917.

Eimas, P. D., Quinn, P. C., & Cowan, P. (1994). Development of exclusivity in perceptually based categories of young infants. *Journal of Experimental Child Psychology*, *58*, 418–431.

Fagen, J. W., & Rovee-Collier, C. (1983). Memory retrieval: A time-locked process in infancy. *Science*, *222*, 1349–1351.

Freedman, D. J., & Assad, J. A. (2006). Experience-dependent representation of visual categories in parietal cortex. *Nature*, *443*, 85–88.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science*, *291*, 312–316.

French, R. M., Mermillod, M., Quinn, P. C., & Mareschal, D. (2001). Reversing category exclusivities in infant perceptual categorization: Simulation and data. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the twenty-third annual conference of the Cognitive Science Society* (pp. 307–312). London: Erlbaum.

Furrer, S. D., & Younger, B. A. (2008). The impact of specific prior experiences on infants' extension of animal properties. *Developmental Science*, *11*, 712–721.

Gelman, R., Durgin, F., & Kaufman, L. (1995). Distinguishing between animates and inanimates: Not by motion alone. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition—A multidisciplinary debate* (pp. 150–184). Oxford, UK: Clarendon Press.

Gelman, S. A. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology*, *20*, 65–95.

Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition*, *65*, 263–297.

Goldstone, R. L., Steyvers, M., & Larimer, K. (1996). Categorical perception of novel dimensions. In G. W. Cottrell (Ed.), *Proceedings of the 18th annual conference of the Cognitive Science Society* (pp. 243–248). Hillsdale, NJ: Erlbaum.

Graham, S. A., Kilbreath, C. S., & Welder, A. N. (2004). Thirteen-month-olds rely on shared labels and shape similarity for inductive inferences. *Child Development*, *75*, 409–427.

Greco, C., Hayne, H., & Rovee-Collier, C. (1990). Roles of function, reminding, and variability in categorization by 3-month-old infants. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 617–633.

Harnad, S. (Ed.) (1987). *Categorical perception: The groundwork of cognition*. New York: Cambridge University Press.

Hayne, H., Rovee-Collier, C., & Perris, E. E. (1987). Categorization and memory retrieval by three-month-olds. *Child Development*, *58*, 750–767.

Horst, J. S., Oakes, L. M., & Madole, K. L. (2005). What does it look like and what can it do? Category structure influences how infant categorize. *Child Development*, *76*, 614–631.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Reviews*, *99*, 480–517.

Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science*, *310*, 863–866.

Iachini, T., Borghi, A. M., & Senese, V. P. (2008). Categorization and sensorimotor interaction with objects. *Brain and Cognition*, *67*, 31–43.

Jiang, X., Bradley, E., Rini, R. A., Zeffiro, T., VanMeter, J., & Riesenhuber, M. (2007). Categorization training results in shape- and category-selective human neural plasticity. *Neuron*, *53*, 891–903.

Kalish, C. W., & Gelman, S. A. (1992). On wooden pillows: Multiple classification and children's category-based inductions. *Child Development*, *63*, 1536–1557.

Kelly, D. J., Liu, S., Ge, L., Quinn, P. C., Slater, A. M., Lee, K., et al. (2007). Cross-race preferences for same-race faces extend beyond the African versus Caucasian contrast in 3-month-old infants. *Infancy*, *11*, 87–95.

Kelly, D. J., Quinn, P., Slater, A. M., Lee, K., Gibson, A., Smith, M., et al. (2005). Three-month-olds, but not newborns, prefer own-race faces. *Developmental Science*, *8*, 31–36.

Kovack-Lesh, K., Horst, J., & Oakes, L. M. (2005, October). *Previous experience influences infants' category learning in the laboratory*. Poster presented at the biennial meeting of the Cognitive Development Society, San Diego, CA.

Kovack-Lesh, K. A., Horst, J. S., & Oakes, L. M. (2008). The cat is out of the bag: The joint influence of previous experience and looking behavior on infant categorization. *Infancy*, *13*, 285–307.

Lin, E. L., & Murphy, G. L. (2001). Thematic relations in adults' concepts. *Journal of Experimental Psychology: General*, *130*, 3–28.

Madole, K., & Cohen, L. (1995). The role of object parts in infants' attention to form- function correlations. *Developmental Psychology*, *31*, 637–648.

Mandler, J. M. (1998). Representation. In W. Damon (Series Ed.) & D. Kuhn & R. S. Siegler (Vol. Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (5th ed., pp. 255–308). New York: Wiley.

Mandler, J. M. (2000). Perceptual and conceptual processes in infancy. *Journal of Cognition and Development*, *1*, 3–36.

Mandler, J. M. (2004). *The foundations of the mind: Origins of conceptual thought*. New York: Oxford University Press.

Mandler, J. M., & McDonough, L. (1993). Concept formation in infancy. *Cognitive Development*, *8*, 291–318.

Mareschal, D., French, R. M., & Quinn, P. (2000). A connectionist account of asymmetric category learning in infancy. *Developmental Psychology*, *36*, 635–645.

Mareschal, D., & Quinn, P. C. (2001). Categorization in infancy. *Trends in Cognitive Sciences*, *5*, 443–450.

Mash, C. (2006). Multidimensional shape similarity in the development of visual object classification. *Journal of Experimental Child Psychology*, *95*, 128–152.

McDonough, L., & Mandler, J. M. (1998). Inductive generalization in 9- and 11-month-olds. *Developmental Science*, *1*, 227–232.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207–238.

Needham, A., & Baillargeon, R. (1998). Effects of prior experience on 4.5-month-old infants' object segregation. *Infant Behavior and Development*, *21*, 1–24.

Needham, A., Cantlon, J. F., & Ormsbee Holley, S. M. (2006). Infants' use of category knowledge and object attributes when segregating objects at 8.5 months of age. *Cognitive Psychology*, *53*, 345–360.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 104–114.

Oakes, L. M., Plumert, J. M., Lansink, J. M., & Merryman, J. D. (1996). Evidence for task-dependent categorization in infancy. *Infant Behavior and Development*, *19*, 425–440.

Oakes, L. M., & Tellinghuisen, D. J. (1994). Examining in infancy: Does it reflect active processing? *Developmental Psychology*, *30*, 748–756.

Pauen, S. (2002). Evidence for knowledge-based category discrimination in infancy. *Child Development*, *73*, 1016–1033.

Proffitt, J. B., Coley, J. D., & Medin, D. L. (2000). Expertise and category-based induction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 811–828.

Quinn, P. C. (2002). Beyond prototypes: Asymmetries in infant categorization and what they teach us about the mechanisms guiding early knowledge acquisition. In R. V. Kail & H. W. Reese (Eds.), *Advances in child development and behavior* (Vol. 29, pp. 161–193). San Diego, CA: Academic Press.

Quinn, P. C. (2004). Development of subordinate-level categorization in 3- to 7-month-old infants. *Child Development*, *75*, 886–889.

Quinn, P. C. (2005). Young infants' categorization of humans versus nonhuman animals: Roles for knowledge access and perceptual process. In L. Gershkoff-Stowe & D. H. Rakison (Eds.), *Building object categories in developmental time* (pp. 107–130). Mahwah, NJ: Erlbaum.

Quinn, P. C., & Eimas, P. D. (1998). Evidence for a global categorical representation of humans by young infants. *Journal of Experimental Child Psychology*, *69*, 151–174.

Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. *Perception*, *22*, 463–475.

Quinn, P. C., Westerlund, A., & Nelson, C. A. (2006). Neural markers of categorization in 6-month-old infants. *Psychological Science*, *17*, 59–66.

Quinn, P. C., Yahr, J., Kuhn, A., Slater, A. M., & Pascalis, O. (2002). Representation of the gender of human faces by infants: A preference for female. *Perception*, *31*, 1109–1121.

Rakison, D. H. (2000). When a rose is just a rose: The illusion of taxonomies in infants' categorization. *Infancy*, *1*, 77–90.

Rakison, D. H. (2005). Developing knowledge of motion properties in infancy. *Cognition, 96,* 183–214.

Rakison, D. H., & Butterworth, G. (1998). Infants' use of object parts in early categorization. *Developmental Psychology, 34,* 49–62.

Rakison, D., & Cohen, L. (1999). Infants' use of functional parts in basic-like categorization. *Developmental Science, 2,* 423–431.

Rakison, D. H., & Hahn, E. R. (2004). The mechanisms of early categorization and induction: Smart or dumb infants? In R. Kail (Ed.), *Advances in child development and behavior* (Vol. 32, pp. 281–322). New York: Academic Press.

Rakison, D. H., & Oakes, L. M. (Eds.) (2003). *Early category and concept development: Making sense of the blooming, buzzing confusion.* Oxford, UK: Oxford University Press.

Rakison, D. H., & Poulin-Dubois, D. (2001). The developmental origin of the animate-inanimate distinction. *Psychological Bulletin, 127,* 209–228.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104,* 192–233.

Rovee-Collier, C. K., Sullivan, M. W., Enright, M., Lucas, D., & Fagen, J. W. (1980). Reactivation of infant memory. *Science, 208,* 1159–1161.

Ruff, H. A. (1986). Components of attention during infants' manipulative exploration. *Child Development, 57,* 105–114.

Saffran, J. R., Pollak, S. D., Seibel, R. L., & Shkolnik, A. (2007). Dog is a dog is a dog: Infant rule learning is not specific to language. *Cognition, 105,* 669–680.

Saiki, J., & Hummel, J. E. (1996). Attribute conjunctions and the part configuration advantage in object category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1002–1019.

Schwarzlose, R. F., Swisher, J. D., Dang, S., & Kanwisher, N. (2008). The distribution of category and location information across object-selective regions in human visual cortex. *Proceedings of the National Academy of Sciences of the United States of America, 105,* 4447–4452.

Schyns, P. G., Goldstone, R. L., & Thibaut, J. (1998). The development of features in object concepts. *Behavioral and Brain Sciences, 21,* 1–17.

Shepp, B. E. (1978). From perceived similarity to dimensional structure: A new hypothesis about perceptual development. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 135–167). Hillsdale, NJ: Erlbaum.

Smith, L. B. (1989). A model of perceptual classification in children and adults. *Psychological Review, 96,* 125–144.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts.* Cambridge, MA: Harvard University Press.

Spear, N. E. (1973). Retrieval of memory in animals. *Psychological Review, 80,* 163–194.

Spear, N. E. (1978). *The processing of memories: Forgetting and retention.* Hillsdale, NJ: Erlbaum.

Sullivan, M. W. (1982). Reactivation: Priming forgotten memories in human infants. *Child Development, 57,* 100–104.

Träuble, B., & Pauen, S. (2007). The role of functional information for infant categorization. *Cognition, 105,* 362–379.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,* 352–373.

Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General, 113,* 169–193.

Van de Walle, G. (1999, April). *Conceptual categorization in infancy: Membership in colour vs. kind categories.* Poster presented at the meeting of the Society for Research in Child Development, Albuquerque, NM.

Welder, A. N., & Graham, S. A. (2001). The influences of shape similarity and shared labels on infants' inductive inferences about nonobvious object properties. *Child Development, 72,* 1653–1673.

Younger, B. A. (1985). The segregation of items into categories by ten-month-old infants. *Child Development, 56,* 1574–1583.

Younger, B. (1990). Infant categorization: Memory for category-level and specific item information. *Journal of Experimental Child Psychology, 50,* 131–155.

Younger, B. A. (1993). Understanding category members as "the same sort of thing": Explicit categorization in ten-month infants. *Child Development, 64,* 309–320.

Younger, B. A., & Cohen, L. B. (1983). Infant perception of correlations among attributes. *Child Development, 54,* 858–867.

Younger, B. A., & Cohen, L. B. (1986). Developmental change in infants' perception of correlations among attributes. *Child Development, 57,* 803–815.

Younger, B. A., & Fearing, D. D. (2000). A global-to-basic trend in early categorization: Evidence from a dual-category habituation task. *Infancy, 1,* 47–58.

Child Development, May/June 2010, Volume 81, Number 3, Pages 898–912

Mexican-Heritage Children's Attention and Learning From Interactions Directed to Others

Katie G. Silva, Maricela Correa-Chávez, and Barbara Rogoff
University of California, Santa Cruz

The study builds on ethnographic research noting an emphasis in many Indigenous communities of the Americas on learning through keen observation of and participation in ongoing community activities. Forty-four U.S. Mexican-heritage 5- to 11-year-old children whose families likely have experience with Indigenous ways more frequently attended to and learned from a toy construction activity that was directed to another child, compared to 36 U.S. Mexican-heritage children whose mothers had extensive experience with Western school (and related European American practices). The results support the idea that children whose family history emanates from Indigenous communities of México may be especially oriented to learning by observing ongoing events, and that this method of learning may be less commonly used by children whose families have extensive experience with schooling (and related Western practices).

This study investigates cultural variation in children's attentiveness to a nearby toy construction demonstration that was not directed to them. We examined this "third-party attention" and resultant learning of U.S. Mexican-heritage children whose families likely varied in experience with Indigenous practices and with Western schooling (and related practices).

Third-party attention can provide children with important information regarding, for example, their community's cultural values and how to interpret ambiguous situations (Lewis & Feiring, 1981; Mitchell-Kernan & Kernan, 1975; Repacholi & Meltzoff, 2007). They can learn new words and the appropriate use of language through listening in on conversations not directed to them (Akhtar, 2005; Barton & Tomasello, 1991; Ochs, 1988; Oshima-Takane, Goodz, & Deverensky, 1996; Schieffelin, 1991; Ward, 1971).

Although keen observation of events directed to others may be helpful to everyone, it may be more common among people with experience of Indigenous practices of the Americas than among people whose experience heavily involves Western schooling and related practices. Ethnographic research in Indigenous communities of the Americas shows that children are integrated in a wide range of their community's activities and encouraged to keenly observe the events around them (Cancian, 1964; Corona & Pérez, 2005; de Haan, 1999; de Leon, 2000; Gaskins, 2000; Greenfield, 2004; Modiano, 1973; Morelli, Rogoff, & Angelillo, 2003; Whiting & Edwards, 1988). For example, Chamoux (1992) reported that Indigenous children in Central Mexico gain a great deal of knowledge through keen observation of ongoing events and understanding of Indigenous explanatory theory by listening in on conversations among adults.

The cultural learning tradition in which children observe and pitch in to ongoing events of importance in their community has been labeled "learning through *intent community participation*" (Rogoff, Paradise, Mejía Arauz, Correa-Chávez, & Angelillo, 2003; Rogoff et al., 2007). As "legitimate peripheral participants" in community events (Lave & Wenger, 1991), children can observe and contribute when they are ready. Indeed, they may not only be given the opportunity to learn by observing ongoing events but they may be expected to spontaneously

---

We greatly appreciate the participation of the children, and the permission of their parents and teachers from Gault School and Starlight School. For able assistance, we thank Mayra Moreno, Anna Corwin, Irma Núñez, Isaí Baltézar, and Carly Bliman. We appreciate the helpful comments of Nameera Akhtar, Maureen Callanan, Amy Dexter Roberts, and Megan Luce on this manuscript. We are grateful for support to KGS from the Center for Informal Learning and Schools (NSF Grant ESI-0119787) and from NIH Graduate Trainee Fellowship T32 HD046423, and for support to M.C.C. from an AERA/IES post-doctoral fellowship. We are also grateful for support from the UCSC Foundation Professorship in Psychology to B.R. Maricela Correa-Chávez's work on this study occurred at University of California, Santa Cruz and at University of California, Los Angeles; she is now at Clark University.

Correspondence concerning this article should be addressed to Katie Silva, 277 Social Sciences 2, University of California, Santa Cruz, CA 95064. Electronic mail may be sent to ksilva@ucsc.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8103-0017

do so throughout their day (Chavajay, 1993; Philips, 1972).

Learning through intent community participation may be present everywhere but less common in communities in which children are seldom included in a wide range of community activities. In middle-class communities where schooling is prevalent, children are commonly segregated from the range of activities of their communities. Much of their days are spent in school and they are otherwise restricted from the work and mature social activities of their communities (LeVine & White, 1992; Morelli et al., 2003; Rogoff, 2003; Zelizer, 1992). As such, they have limited opportunities to learn by watching or listening to events that are not directed to them. Instead, adult-child interactions often involve child-centered lessons directed to children, utilizing formats common in Western schooling (Harkness, 1977; Heath, 1983; Martini, 1996; Ochs & Schieffelin, 1984; Rogoff, Mistry, Göncü, & Mosier, 1993). Frequently, middle-class parents attempt to manage their children's attention, unlike the expectation in some Indigenous-heritage communities where children are expected to attend keenly on their own (Chavajay, 1993; Dixon, LeVine, Richman, & Brazelton, 1984; Rogoff et al., 1993).

Systematic Cultural Comparisons of Observation and Attention

Although ethnographic studies often have implied a contrast with Western middle-class ways of teaching and learning, systematic cultural comparisons of patterns of attention and observation are rare. One comparative study examined attention when children were directly instructed and found that U.S. Mexican-heritage children from Indigenous regions of México (whose mothers had only basic schooling) more frequently observed an origami demonstration and their peers' efforts to fold the figures, without pressing for more information, compared with both European heritage and Mexican-heritage U.S. children whose mothers had extensive school experience (Mejía Arauz, Rogoff, & Paradise, 2005). The authors related this difference to community emphasis on learning through observation versus experience with the lesson formats of schooling and other Western practices.

Several studies have found cultural differences in simultaneity of attention to multiple surrounding events. Children whose families came to the United States from Indigenous regions of México often skillfully attended to several events simultaneously, without interrupting their attention to one event as they attended to another. They did so more often than European American middle-class children and U.S. Mexican-heritage children whose mothers had extensive school experience, who more often attended to one event at a time, interrupting each in rapid alternation (Correa-Chávez, Rogoff, & Mejía Arauz, 2005). Similarly, Guatemalan Maya toddlers and their mothers more often simultaneously attended to several events at once compared to middle-class European heritage toddlers and mothers, who tended to alternate their attention between events (Chavajay & Rogoff, 1999; Rogoff et al., 1993). These authors speculated that attending simultaneously to several events would facilitate learning by observing, in that it would allow monitoring of possibly important surrounding events without interrupting the activity at hand.

To our knowledge only two studies have systematically examined cultural variation in children's attention to surrounding events that were *not* directed to them. Spontaneously attending to surrounding events can be key to learning in communities where children have opportunities to be present during the range of community interactions. The two studies, with children living in two Indigenous communities (Navajo and Guatemalan Maya), found differences in children's attention to third-party interactions that may relate to experience with Indigenous ways as well as to experience with Western schooling.

In one study, children were on the periphery as one child taught a game to another. Middle-class European American children were more often off task and distracted than Navajo children, who more often maintained attention to the game and the instruction (Ellis & Gauvain, 1992).

In the other study, children were present when an adult showed their sibling how to make a toy. Guatemalan Maya children from families with many traditional Maya practices (and little or no Western schooling) attended more keenly to the interaction that was directed to their sibling than did middle-class European American children and Guatemalan Maya children whose mothers had extensive experience with Western schooling (Correa-Chávez & Rogoff, 2009). The Maya children from more traditional families also demonstrated greater learning—they needed less assistance when they unexpectedly had the chance to make the toy they had had the opportunity to observe being made a week earlier.

Both studies are consistent with ethnographic work showing that learning through keen observation to events that involve others is important in Indigenous communities of the Americas. Both of the prior studies examined Indigenous children's attention in communities that are structured as and identify as Indigenous (Navajo in the United States and Mayan in Guatemala).

The present study examined the attention of children whose families were *not* living in Indigenous communities but whose families are likely to have experience with practices emanating from Indigenous communities of prior generations in México, although they currently live in a predominantly European American setting, with a distinct structure of learning opportunities. We examined whether similar patterns of attention would be seen among children in families that had emigrated to the United States from rural regions of México that formerly had extensive Indigenous populations and who were likely to have experience with some Indigenous practices, compared with children from families from the same region of México but with extensive experience in Western schooling and related practices.

Thus, although much cultural research treats national origins as homogeneous, our study focuses on cultural practices that vary within two groups with the same national origin (Mexican) and the same residence (California). Our focus is on experience with cultural ways, rather than on ethnic identity per se.

The present study also differs from the prior two studies in extending to the practices of distinct Indigenous groups—those of West Central México rather than of the United States and Guatemala. (The Indigenous population of the region of México from which most of our sample emigrated has been largely P'urepecha in addition to a few smaller groups—all of which are culturally and linguistically very distinct from both Navajo and Mayan.)

The practices of Indigenous communities and of Western schooling appear to be widespread features of constellations of cultural practices that are important for many U.S. immigrant children, as well as U.S. and Canadian Native children and children in many communities of Mexico and Central America. However, the generality across these national and cultural groups needs to be investigated, not assumed. (We limit our speculations regarding generality of the patterns to North and Central America in the interest of caution, given that we have less knowledge of the extent to which the pattern appears on other continents.)

Identifying Experience With Indigenous Practices of the Americas and Western Schooling

A large portion of the Mexican immigrant population in central California has come from rural areas in México that have been considered Indigenous in the recent past, such as in Michoacán. People of some rural *pueblos* in this region of México consider themselves to be Indigenous. Many others consider themselves *mestizos* (mixed race), but in earlier generations many of their ancestors lived as Indigenous people in *pueblos* that were often organized according to recognizably Indigenous practices (Bonfil Batalla, 1996; Frye, 1996; Vigil, 1998). These pueblos have had little access to Western schooling. Therefore, we presume that experience with Indigenous practices is likely among families that have immigrated to the United States from such regions of México if they have limited schooling.

Our presumption of historical connection with Indigenous practices is supported by a study in a West Central Mexican town that was considered Indigenous a generation ago. Residents of this town have little involvement in Western schooling and participate in some Indigenous practices that can be traced back centuries; highly schooled residents of nearby Guadalajara who have lived in the city for only a generation engage in Indigenous practices in an attenuated form, whereas families with generations of extensive schooling and residence in Guadalajara do not engage in these Indigenous heritage practices (Najafi, Mejía Arauz, & Rogoff, 2007).

Similarly, in a Nahua town in Central México, people usually do not identify themselves as Indigenous (except to claim or request something from the government). They may simply refer to themselves as "Mexicans," although they refer to people in nearby towns as Indigenous. However, they continue to pass down key traditional Mesoamerican Indigenous practices and beliefs across generations, including stories, moral values, and cosmological concepts (Lorente y Fernández, 2006; personal communication, June 2007).

Luis Urrieta (2003) reflected on the Indigenous identity of people from the West Central Mexican region from which many California immigrant families come. His father's *pueblo* in Michoacán had once been "a culturally *P'urhépecha* village. . . . But due to the proximity of the *hacienda* there were many cultural shifts that diluted the strength of Indigenous identity" (p. 155). In his youth, Urrieta found references to being Indian confusing:

> I heard older relatives in Los Angeles and Michoacán saying, "When we were Indians," while at other times, "When we were more Indian." This was also made more confusing when referring to people of nearby *pueblos* in the Pátzcuaro region as being *mas indios*, or "more Indian," as if being "Indian" was something that could be diluted, lessened, or changed. (p. 149)

Indigenous identity underwent some changes around the time of the Mexican Revolution in the early 1900s, when many Indigenous *pueblos* transformed from Indian to *campesino*, in a process promoted in part by schooling (Urrieta, 2003). Only recently has extensive schooling become widespread, extending from the United States and Europe to Indigenous communities in Mesoamerica (Meyer, Ramirez, & Soysal, 1992). One of the goals of mass schooling in México was to "modernize" the mostly Indigenous countryside, and to suppress Indigenous ways (Bonfil Batalla, 1988; Stavenhagen, 1988).

With experience in particular cultural institutions, such as school, people may adopt approaches resembling practices associated with those institutions in other parts of daily life (Gutiérrez & Rogoff, 2003). Mothers' experience with schooling in such communities often relates to their use of teaching approaches that coincide with teaching models often used in Western schooling (Chavajay, 2006; Chavajay & Rogoff, 2002; Dunn, Griggs, & Price, 1993; Laosa, 1980, 1982; Moreno, 2000; Rogoff et al., 1993). Mexican-heritage U.S. children whose mothers had completed at least 12 grades observed an Origami demonstration in ways that resembled those of U.S. European-heritage children whose mothers had also completed at least 12 grades. Both backgrounds relied less on observation as a source of information and less often attended simultaneously to ongoing events compared with Mexican-heritage U.S. children whose mothers averaged six grades of school (Correa-Chávez et al., 2005; Mejía Arauz et al., 2005).

Although we argue that experiences with school practices provide familiarity with particular ways of teaching and learning (along with LeVine, LeVine, & Schnell, 2001), we do not regard schooling as a solitary "active ingredient" in community differences. Rather we regard the extent of experience with Western schooling in Indigenous-heritage communities of México and Guatemala as part of a constellation of associated characteristics—such as migration, urban experience, occupations requiring school credentials, few children in a family, and limited involvement of extended family—that may also relate to teaching and learning approaches (Chavajay & Rogoff, 2002; Rogoff & Angelillo, 2002; Rogoff, Correa-Chávez, & Navichoc Cotuc, 2005; Rogoff et al., 2007). Thus, family experience with Western schooling may contribute, along with many associated practices, to community practices surrounding learning.

The Present Study

Our primary question was whether U.S. Mexican-heritage children whose families likely had greater familiarity with Indigenous practices (and little experience of Western schooling) would use more sustained attention while their sibling is shown how to construct a toy, compared with U.S. Mexican-heritage children whose mothers had extensive experience with school. In addition, we examined the children's learning, by coding the ease with which the children of the two backgrounds constructed the toy their sibling had been shown how to make, when they were given an unexpected opportunity to make it a week later. We also examined whether the children's extent of sustained attention to the third-party interaction in the first session was related to their subsequent ease of constructing the toy that they had had the opportunity to observe.

The current study also examined differences in children's attempts to direct attention toward themselves when they were not included in the interaction. Children who are accustomed to being directly addressed may disrupt interactions that are addressed to others by calling attention to themselves. European American middle-class children more often interrupted and sought attention when adults were occupied than did Efe or Maya children (Correa-Chávez & Rogoff, 2009; Rogoff et al., 1993; Verhoef, Morelli, & Anderson, 1999). In ethnographic reports, children in several Indigenous Mexican communities rarely tried to garner attention to themselves (de Haan, 1999; Gaskins, 2000; Paradise, 1996). Both Guatemalan Mayan children from traditional families and those whose families had extensive experience with Western schooling seldom tried to divert attention to themselves when they were not addressed (Correa-Chávez & Rogoff, 2009).

Method

Participants and Their Communities

Eighty children between the ages of 5 and 11 years old participated in the study (see Table 1

Table 1
Maternal Schooling, Gender, and Age of Focal Child

	Pueblo Basic Schooling (*n* = 22 pairs)	Mexican Hi Schooling (*n* = 18 pairs)
Average maternal schooling	6.4 grades (range = 0–9)	12.6 grades (range = 12–16)
First nonaddressed child (observing mouse construction)		
Gender	7 girls, 15 boys	11 girls, 7 boys
Average age of focal child	6.8 years (range = 5–10 years)	6.4 years (range = 5–9 years)
Second nonaddressed child (observing frog construction)		
Gender	9 girls, 13 boys	11 girls, 7 boys
Average age of focal child	9.0 years (range = 6–11 years)	9.1 years (range = 7–10 years)

for ages, gender, and grades in school for children). Children were Mexican-heritage sibling pairs recruited through two public elementary schools in neighboring California towns where immigrant populations from México have often come from rural areas of the state of Michoacán, attracted by the agricultural industry. Siblings were used to ensure that the children working together had similar family schooling histories, to provide a familiar partner, and to provide a reasonable excuse for making the children ''wait'' for another child.

In the first of the two sessions, participants were 44 children from ''Pueblo Basic Schooling'' families that likely had familiarity with Indigenous practices from the pueblos of México and had little experience with Western schooling (average maternal grade = 6, range = 0–9 grades) and 36 children from ''Mexican Hi Schooling'' families (average maternal grade = 13, range = 12–16 grades). In the second session, 37 of the children whose families were from the Pueblo Basic Schooling background and 35 of the children from Mexican Hi Schooling families participated. (Some children who participated in the first session were not able to participate in the second session 1 week later due to absence from school.)

Maternal experience with school was divided at high school completion because completion of high school is an important milestone in both México and the United States that can often lead to different occupations and income, and often accompanies differences in family life such as fewer children and more school-like interactions of parents with children. Maternal schooling experience was thus a proxy for a constellation of cultural practices that seem to be indicative of Indigenous experience for families who come from rural areas of México to California for agricultural work. The labels ''Pueblo Basic Schooling'' and ''Mexican High Schooling'' represent not only differences in extent of the children's mothers' experience in Western schooling but also many other related features such as their fathers' and grandparents' schooling, parental occupations, and family size.

Of the 44 children from Pueblo Basic Schooling families, most (32) were born in the United States, 6 were born in México, and parents declined to give birthplace information of 6 of the children. Almost all (21) of the 22 Pueblo Basic Schooling mothers were born in México and 1 was born in California; almost all (20) had attended school in México, 1 had attended school in both México and the United States, and 1 had not attended school. Of the fathers, almost all (20) were born in México and 2 of their birthplaces were not reported; they averaged 7 grades of schooling. In both backgrounds, mothers' and fathers' schooling was highly correlated, $r = .66$, $p < .01$. Grandparents' schooling was limited to primary grades: Most (65%) had completed 0–3 grades, and some (23%) had completed 4–6 grades; information was not available on the remainder. The Pueblo Basic Schooling families averaged 3.3 children.

Of the 36 children from Mexican Hi Schooling families, almost all (33) were born in the United States, 1 was born in México, and parents declined to give information about the birthplace of 2 children. Half (9) of the 18 Mexican Hi Schooling mothers were born in the United States and half (9) were born in México; almost all (16) had attended school in the United States, 1 had attended school in México, and 1 had attended school in both countries. Of the fathers, about half (7) were born in the United States, 8 were born in México, and 3 of their birthplaces were not reported; they averaged 13 grades of schooling. Grandparents' schooling ranged broadly: 0–3 grades completed (33%), 4–6 grades (11%), 7–9 grades (6%), 10–11 grades (11%), and 12 or more grades (18%). The Mexican Hi Schooling families averaged 2.4 children.

Procedure

One of two bilingual female research assistants ("Toy Ladies") who were blind to the hypotheses and goals of the study worked with the siblings in a vacant section of an elementary school classroom. Both Toy Ladies worked with children from both groups and were not informed of the background of the children with whom they were working. Comparison of sessions from the two Toy Ladies ensured that differences between Mexican Hi Schooling and Pueblo Basic Schooling groups were not due to the different Toy Ladies who worked in different schools.

The procedure was based on Correa-Chávez and Rogoff's (2009) study with Mayan and European American children, for comparability across the different cultural communities. In Session 1, each sibling had a chance to make a toy; our interest was in each child's attention when they were not addressed. The first nonaddressed child was always the younger child (waiting while the sibling made a toy mouse that runs via a spool and rubber band) and the second nonaddressed child was always the elder (waiting while the sibling folded an Origami paper frog that jumps). Thus, the children's age, order of making a toy, and which toy they had a chance to observe being constructed were purposely not counterbalanced. Hence we do not compare age, order, or toy differences. We note, however, that the pattern of the younger child who observed the mouse construction first was similar to that of the older child who observed the frog construction second. In Session 2, the children came back individually 1 week later to pick up the toys they had made, and were unexpectedly invited to construct the toy that their sibling made the week before.

Session 1: Observing sibling's toy construction. The Toy Lady led the sibling pairs into a child-friendly classroom with two tables. The main table had a chair for the Toy Lady and one for the child constructing the toy. A side table, 16 in. away, had a chair for the nonaddressed sibling to sit in while "waiting." The nonaddressed child's chair was parallel to and slightly forward of the sibling's chair, to give a cue that the nonaddressed child was not expected to participate in the activity of their sibling. The position required the nonaddressed child to turn to watch, which also meant that coders could clearly see when the child was attending to the construction activity (see Figure 1). All interactions were videotaped.

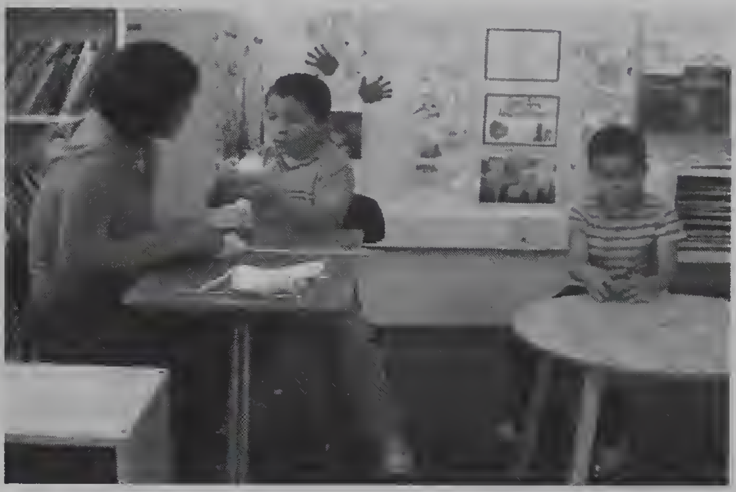


Figure 1. Setup for Session 1.
Note. The Toy Lady and a child make the mouse at the rectangular table while the nonaddressed sibling sits at the round table with the distracter.

The session was designed for the nonaddressed child to know that they would make a different toy than their sibling and thus would be unlikely to think that attending to their sibling's mouse construction would help for making their own toy. The Toy Lady greeted the children and introduced herself and told them that she was going to help them make a toy. She led the younger child to wait in the seat at the side table saying, "I'm going to start with your big brother/sister, and he/she is going to make a mouse, and when he/she's done you can make a frog" (referring to the models of each toy in view on the main table). The Toy Lady handed the nonaddressed sibling the distracter toy saying, "While he/she does that, how about you sit here and you can play with this." The distracter toy was a "do-nothing" machine—a wooden block with a crank to turn. This toy was chosen because it does not require much focused attention to manipulate it and it is attractive but is not interesting for long.

After settling the nonaddressed child, the Toy Lady then sat with the older sibling at the main table and put two sets of the materials used to make the mouse toy on the table (one set for the child and one for herself). She addressed only the child making the mouse: "I'm going to be making one and you are going to be making one. You can watch me and do what I do."

After the mouse was completed, the Toy Lady asked the siblings to switch seats. The child who constructed the mouse now sat at the nearby table and was given the distracter toy and told, "You can play with this while you wait if you want." The previously nonaddressed child sat at the table with the Toy Lady and constructed an origami jumping frog. The Toy Lady said, "Okay (addressed child), now it's your turn." The Toy Lady followed a script similar to the one used while constructing the mouse.

At the end of the session, the Toy Lady had the children put their names on a bag for her to keep their toys in for a week (to prevent other children in the school from studying them and to justify Session 2). She told the children "We are going to keep your toys for a few days until all of the kids have a chance to make one. You can come back next week to pick them up." Session 1 lasted about 8 min in total for both children to construct their toy.

The Toy Lady followed a script in all interactions to ensure that she taught and interacted with all children in a similar way. The script limited the Toy Lady's extent of explanation of steps in the toy construction (avoiding explicit verbal instructions and instead saying, e.g., "put this through here") to ensure that the nonaddressed child would need to watch the toy construction in order to learn from it.

The script also provided the Toy Lady with guidelines on how to react if the nonaddressed sibling sought involvement and/or attempted to enter into the construction of the toy: She gently discouraged the nonaddressed sibling's participation by reminding the child that she/he would soon have the chance to make the origami frog. The script also included instructions for how to interact if the nonaddressed child tried to disrupt the construction activity. For example, if the nonaddressed child began to talk about something extraneous such as an upcoming vacation, the Toy Lady would either ignore the child or say, "oh that's nice" and turn back to the child she was working with and begin the next step, making it clear that she was only addressing the child at the table.

The Toy Lady was not told about the second session until she had completed Session 1 with all of the children, to further limit the possibility of her encouraging attention from the nonaddressed child. The Toy Lady was not told about our interest in the nonaddressed child's attention (or about our interest in the children's backgrounds). She was informed that we were interested in how children learn to make toys and what the nonaddressed children do naturally when waiting for their turn to make a toy.

A procedural check was conducted on 25% of the Session 1 data in order to ensure that the Toy Lady was consistent in her treatment of the nonaddressed child and her adherence to the script. Most important, the Toy Ladies did not address the nonaddressed child or encourage that child to attend to the toy construction in any sessions; they interacted almost exclusively with the child constructing the toy.

Session 2: Nonaddressed siblings' learning. Each child returned to the classroom individually, approximately 7 days later, to pick up the toy that they made a week before. The Toy Lady unexpectedly gave them the opportunity to make the toy that their sibling had made the previous week, saying, "I have some extra material and I thought that you might like to make a [frog/mouse] this time."

The Toy Lady put all of the materials that the child needed on the table and said, "Here are all the materials that you will need to make a [frog/mouse]. I have a project to do right now so you can get started and let me know if you need help." The Toy Lady then picked up a notebook and started writing, seemingly absorbed in her project, so that the child would not expect her to give instructions and so that the Toy Lady would have reason not to respond immediately if the child requested help (see Figure 2). The Toy Lady sat at the same table as the child but with her chair slightly pushed away from the table, working on her project and sitting back in her chair, to create distance from the child and the toy construction.

To measure learning, we examined the extent of help needed by the child to construct the toy, based on a set of graduated, scripted hints. (A similar series of hints was specified for each step in the mouse and the frog.) As she worked on her project, the Toy Lady inconspicuously monitored the progress of the child and at the beginning of each step of the toy construction, the Toy Lady waited approximately 5 s for the child to attempt to get started on

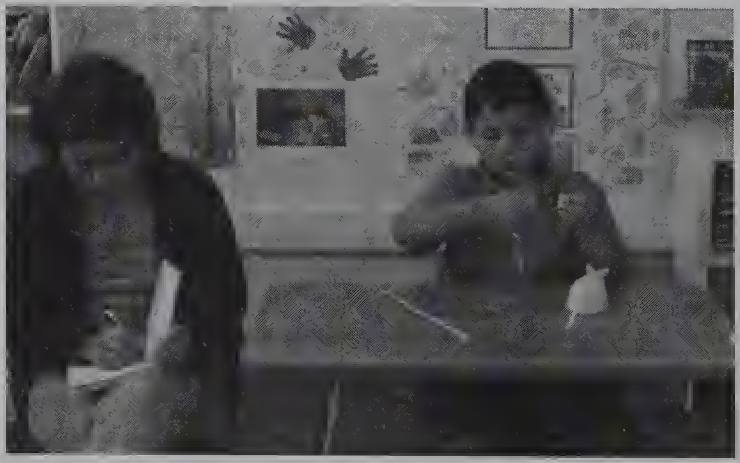


Figure 2. Setup for Session 2.
Note. The Toy Lady acts busy at work on her project while the child attempts to construct the mouse toy.

the step. After the 5 s, if the child did not spontaneously complete the step, or if they asked for help (either verbally or nonverbally) or began to do the step incorrectly, the Toy Lady started off giving them a small hint specifying what step to do (e.g., "The body is first") or what material to begin with (e.g., "Try starting with the string"), then went back to her work and waited another 5 s. If the child continued to have trouble, the Toy Lady gave the child a bigger hint about what to do with the materials (e.g., "Fold this," pointing to the portion of the paper the child needed to fold). If the child needed more help, the Toy Lady demonstrated a bit of the step to the child, and finally if the child could not do the step, the Toy Lady completed the step for them and then went back to her work and told the child to go ahead. A procedural check conducted on 25% of the Session 2 data showed that the Toy Ladies did not differ in adherence to the script, time lapse before giving children hints, or order of hints given.

On completion of the toy, the child was given both toys. (All participants in the same school completed Session 1 before Session 2 began, so that children would not anticipate making their sibling's toy at the time they participated in Session 1.) Each child's second session took an average of 7 min.

Coding of Session 1: Children's Attention to the Construction Activity Plus Disruption and Attempts to Become Involved

Our primary interest was the attention of the nonaddressed child to the construction activity, which was coded in 5-s segments by a coder who was blind to the goals of the study and to the children's maternal schooling. The child's predominant form of attention during each 5-s segment was coded as either sustained attending to the construction, briefly glancing at the construction, or not attending to the construction activity, based on the children's gaze, posture, gestures, and other visual and verbal cues.

Sustained attending to the construction consisted of energetically watching the construction activity, with attention devoted to it for most of the segment. During sustained attending, children sat still and actively attended to the toy construction activity, with their eyes fixed on the construction activity and with an alert body posture. Children could break their attention to the construction momentarily or simultaneously work the distracter toy, but their attention to another activity was minor within the segment.

Glancing involved sporadic interest in the construction activity for a brief portion of a segment, without seemingly being engaged by the construction activity; most of the segment involved attending to something other than the construction activity. In the few instances when both sustained attending and glancing occurred in the same 5-s segment, the segment was counted as sustained attending. (There was no difference between the two backgrounds in the percentage of segments in which children equally used both sustained attending and glancing in the same 5-s segment.)

If children stretched or strained or moved their chair closer to get a better view of the toy construction activity while engaged in sustained attention or glancing, straining to observe was also coded within those 5-s segments.

If the child was not attending to the construction activity, we recorded what else they were attending to. Often, this involved working the distracter toy, but it also sometimes involved looking at objects in the room such as posters and books, or daydreaming.

We also examined the time segments in which the nonaddressed child tried to disrupt the construction activity, trying to divert the attention of the Toy Lady or their sibling to themselves (e.g., by playing loudly in an insistent repetitive way, talking to the sibling or Toy Lady about a topic extraneous to the toy construction, or grabbing materials to get attention). Disruptions were not mutually exclusive with the attentional categories; however, they did not often occur in the same 5-s segment with sustained attention or glancing.

In addition, we coded whether the nonaddressed child showed attempts to collaborate in the construction activity of the Toy Lady and the sibling, verbally and/or nonverbally. Attempts to collaborate included picking up the materials to try to become involved in the construction (not to disrupt it), attempting to help build the toy, nodding to verify the information the Toy Lady was giving about how to make the toy as if they were part of the interaction, asking "Can I do it?" or commenting on the ongoing activity (e.g., "Wow, that looks really cool").

Coding Session 2: Children's Learning: Unexpected Toy Construction

Session 2 was coded by a blind coder for the amount of assistance that the child needed from the Toy Lady in each step of the construction. A different blind coder was used than in Session 1 so that the coder would not know the children's extent of observing in Session 1. The coder tallied the level of hints the child needed to complete each step: completing the step with no assistance (0 points); a small hint about the step or the materials to use for the step (1 point); a bigger hint about what to do with the materials (2 points); showing the child a bit of the step (3 points); or doing the entire step for the child (4 points). The maximum score for the four steps of the mouse was thus 16 and for the five steps of the frog the maximum was 20; no help needed was scored 0.

Reliability. A bilingual coder, blind to the hypotheses of the study, coded the Session 1 videotapes. A second coder coded 40% of the data. Intercoder reliability was calculated by means of Pearson's correlations between the two coders: percent of segments with sustained attention $r = .95$, glances $r = .84$, straining to observe $r = .91$, not attending $r = .99$, disruptions $r = .97$, and attempts to collaborate $r = 99$.

For the Session 2 data, a different bilingual research assistant coded 40% of the data. Intercoder reliability was examined by means of Pearson's correlations between the blind coder and the first author for amount of help needed by the children, $r = .99$. Disagreements in both sessions were resolved by the coders reviewing the segment together and coming to consensus.

Results

We first examine cultural differences in Session 1 in children's attention to an interaction that is directed at others, attempts to disrupt, and attempts to collaborate in others' activities. We then compare the children's learning in Session 2 (i.e., the amount of help they needed to construct the toy that they had the opportunity to observe being constructed a week before). Finally, we examine the relation between attention in Session 1 and learning in Session 2. We employed planned comparisons (a conservative test) for the key analyses for which we had clear predictions.

Attention, Disruption, and Attempts to Collaborate by the Nonaddressed Child (Session 1)

Because the length of sessions varied among individual children, the proportion of time segments was used to analyze the data. However, there was no systematic difference in the length of sessions of the two cultural backgrounds.

Table 2

Mean Percent (and Standard Deviation) of Segments in Which Nonaddressed Children Attended to the Construction

Form of attention	Pueblo Basic Schooling	Mexican Hi Schooling
First nonaddressed child (observing mouse construction)		
Sustained attention	40.2 (19.9)*	28.0 (20.4)*
Not attending	43.8 (19.1)*	53.8 (21.2)*
Brief glancing	16.0 (9.1)	18.1 (11.6)
Second nonaddressed child (observing frog construction)		
Sustained attention	40.5 (32.5)*	19.2 (20.0)*
Not attending	44.6 (24.5)**	62.7 (19.2)**
Brief glancing	14.9 (12.7)	18.0 (11.6)
All siblings combined		
Sustained attention	40.4 (27.0)**	23.6 (20.0)**
Not attending	44.2 (21.7)**	58.3 (20.4)**
Brief glancing	15.4 (10.9)	18.1 (11.0)

*$p < .05$. **$p < .005$.

Third-party attention. Children observing first and those observing second showed the same patterns of attention in both backgrounds, so we focus on the combined data. (Note that any difference between siblings was not interpretable in terms of age, birth order, nature of the toys constructed, or order effects, as these were purposely not counterbalanced.) There were not systematic gender differences in the children's attentional patterns. Table 2 shows the means and standard deviations for Session 1 analyses, separately for each child and then combined.

As expected, U.S. Mexican-heritage children from Pueblo Basic Schooling families used more sustained attention than the U.S. Mexican-heritage children from Mexican Hi Schooling families, in an average of 40% versus 24% of the time segments, $t(78) = 3.18$, $p < .001$ (see Table 2). Straining to observe their sibling's toy construction activity was not common but, as expected, children from Pueblo Basic Schooling families strained to observe more often than did Mexican Hi Schooling children, in an average of 3.9% versus 1.6% (*SDs* = 5.7 and 3.0) of the segments, respectively, $t(78) = 2.39$, $p < .01$.

Also as expected, children from Mexican Hi Schooling families spent more time not attending to the construction compared to children from Pueblo Basic Schooling families, in 58% versus 44% of the time segments, $t(78) = 2.96$, $p < .002$. We had no prediction for differences between the backgrounds in briefly glancing at the construction activity, and analysis showed none, $t(78) = 1.08$, $p = .28$.

When the children were not attending to the construction, they were mostly attending to the distracter toy (in 78% of the nonattention segments). They attended to a made-up activity or game in 16% of the nonattention segments, attended to extraneous activities of the Toy Lady or their sibling in 5% of nonattention segments, and stared into space in 1% of the nonattention segments. There were not differences between the backgrounds in how often their nonattention involved one activity or another.

Nonaddressed sibling's attempts to disrupt or collaborate in the construction. As expected, disruptions were rare among the U.S. Mexican children from both backgrounds, and there was no significant difference between them. Children in the Pueblo Basic Schooling group attempted to disrupt in 2% of the segments, as did children in the Mexican Hi Schooling group (*SD*s = 0.05, 0.04). These rare instances of disruption were in segments in which children were not attending to the toy construction activity, and there were no gender differences. These findings are similar to Correa-Chávez and Rogoff's (2009) findings with two Guatemalan Mayan backgrounds varying in experience with traditional Mayan ways and with Western schooling; they found that disruption was more frequent among European American children.

We expected that if there were differences in attempts to collaborate in the construction activity, children from the Pueblo Basic Schooling background would do so more often. Indeed, the children from Pueblo Basic Schooling families attempted to collaborate in the construction activity more than the children from Mexican Hi Schooling families, in an average of 6.9% of the segments versus 0.7% of the segments (*SD*s = 15.0, 2.2), respectively, $t(78) = 2.70$, $p < .01$. There were not systematic gender differences. The children's attempts to collaborate occurred when they were attending to the activity, suggesting that these attempts were motivated by interest and not by boredom. The greater attempts to collaborate in the construction among nonaddressed children whose families likely had more experience with Indigenous practices than among children whose families had more experience with schooling may relate to observations that Indigenous communities emphasize learning through *intent community participation* (Rogoff et al., 2003; Rogoff et al., 2007), with children pitching in collaboratively with other children and adults in ongoing events.

Learning: Amount of Help Needed in the Unexpected Toy Construction (Session 2)

We present the comparisons separately for the children constructing the mouse and those constructing the frog, because across the two backgrounds, children who constructed the mouse in Session 2 needed less help than children who constructed the frog toy (even though these were the younger siblings), $t(70) = 4.73$, $p < .001$. Much of the information about how to construct the mouse was contained in the materials themselves. For example, even without having watched, a child might assume that the string needed to be wrapped around the spool, and do this without help. It would be more difficult to infer a correct fold for the origami frog. Indeed, Correa-Chávez and Rogoff (2009) found significant differences in amount of help needed only for constructing the frog, among Guatemalan Mayan and European American children.

The scores for amount of help needed were converted to percentages, because the maximum scores differed for the two toys (due to the differing numbers of steps needed to construct them). Planned contrasts were used to test for differences between the two backgrounds in the amount of help needed to construct each toy after having had the opportunity to observe their sibling construct it a week earlier.

The children in the Pueblo Basic Schooling group needed less help than the children in the Mexican Hi Schooling group to construct both the mouse toy and the frog toy: $t(32) = 2.20$, $p < .02$ for the mouse; $t(36) = 1.87$, $p < .04$ for the frog. There were no significant gender differences. Table 3 shows the means and standard deviations for Session 2 analyses.

Children who engaged in more sustained attention to the toy construction in Session 1 needed less help when unexpectedly given the chance to construct their sibling's toy in Session 2. (This was primarily the case for younger children from Pueblo Basic Schooling background constructing the mouse and older children from Mexican Hi Schooling background constructing the frog. It is not clear why the correlations for the other two situations were not significant; we did not expect differences in the extent of correlation across the four cells of Table 3.) Combining across the two cultural backgrounds, $r = -.34$, $p < .02$ for the mouse and $r = -.30$, $p < .04$ for the frog. There were no significant correlations between brief glances and the amount of help needed to construct either toy.

Table 3

Mean Percent (and Standard Deviations) of Help Children Needed to Construct the Toy, and Correlations With Session 1 Sustained Attention

	Pueblo Basic Schooling ($n = 22$ pairs)	Mexican Hi Schooling ($n = 18$ pairs)
First nonaddressed child (observing mouse construction)		
Help needed (% of the available hints)	26.8 (14.8)$_a$	38.6 (16.4)$_a$
Relation of sustained attention and help needed	$r = -.40$*	$r = -.13$
Second nonaddressed child (observing frog construction)		
Help needed (% of the available hints)	46.5 (18.3)$_b$	56.1 (12.6)$_b$
Relation of sustained attention and help needed	$r = -.06$	$r = -.66$**

Note. Subscript letters indicate significant differences across columns with the same letter at *$p < .05$, **$p < .005$.

The cultural differences in how much help the children needed in Session 2 were directly related to the differences in how much sustained attention they showed in Session 1. This was determined by an analysis of covariance (ANCOVA, with cultural background as a fixed factor and amount of sustained attention as a covariate) showing that sustained attention was a significant covariate in how much help the children received, $F(1, 72) = 4.69$, $p < .05$. After controlling for sustained attention, cultural background became a nonsignificant main effect, $F(1, 52) = 2.43$, $p = .12$. These results suggest that other possible cultural differences (such as possible differences in dexterity) did not contribute significantly to the children's learning.

Discussion

This study found cultural differences in children's attention and learning during interactions directed to others in two Mexican-heritage communities living in the United States. The findings extend research indicating that learning through keen attention to surrounding events is emphasized in Indigenous communities of the Americas. The two cultural backgrounds of this study both involved families that had emigrated from regions of México that formerly had extensive Indigenous populations and limited Western schooling. We inferred that the families that have limited parental schooling in the current generation are likely to have experience of Indigenous practices of México, although they do not now live in Indigenous communities. Mexican-heritage families with extensive experience in Western schooling are likely to have less involvement with Indigenous practices, and of course they have more experience with potentially conflicting Western school and related practices.

As predicted, U.S. Mexican-heritage children from families that likely had experience with Indigenous practices and had little experience with formal school attended more keenly to an interaction that was not addressed to them and more often strained to observe it than U.S. Mexican-heritage children whose mothers had extensive experience in Western schooling (and related practices). U.S. Mexican-heritage children whose mothers had extensive experience with school instead often played with the distracter toy, daydreamed, and looked at other items in the room.

The U.S. Mexican-heritage children from families with likely experience of Indigenous practices also learned more about how to construct the toy they had had the opportunity to observe being constructed than the U.S. Mexican-heritage children whose mothers had extensive school (and related) experience. Indeed, sustained attention seemed to have helped the children learn about constructing the toy—children who showed more sustained attention often needed less help when given an unexpected opportunity to construct the toy.

The findings suggest that Indigenous ways of learning may be maintained among families that do not live in Indigenous communities and have left regions where Indigenous practices may be sustained among communities that do not at present identify as Indigenous. Cultural practices may be resilient among populations whose cultural identity is no longer explicitly tied to prior communities from which the practices emanate. In particular, families may continue to structure children's learning in line with the practices of communities that they no longer live in. Our findings regarding attentiveness to surrounding events support the idea that some practices related to the cultural tradition of organizing learning through intent community participation may be sustained among Mexican-heritage families coming from regions where Indigenous communities have been common and Western schooling has been rare.

Generality Across Times and Places

Our study thus suggests some generality in the pattern of cultural differences in children's attention to surrounding events, across populations residing in different nations and with historical or with current ties with Indigenous communities. The families of the present study, who likely have some experience with Indigenous practices of México but who live as newcomers in a largely European American region, not in Indigenous communities, show a pattern similar to that found in two comparative studies performed in Indigenous communities (Correa-Chávez & Rogoff, 2009; Ellis & Gauvain, 1992). Taken together, the three studies support the idea that children from families that have experience with Indigenous practices of the Americas, whether or not they live in or identify with Indigenous communities, are more likely to be alert to and learn from surrounding events than children from families with little or no experience of such practices. In addition, our study extends the pattern to a different Indigenous heritage, involving P'urepecha and other groups of West Central México, which are culturally and linguistically quite distinct from both the Maya of Guatemala and the Navajo of the United States.

Although we feel confident assuming that maternal schooling experience can be a proxy indicator of Indigenous experience in our participant communities given their histories, in this study we did not have specific information on family experience with Indigenous ways. In future research with immigrants from México, it will be helpful to examine experience with Indigenous practices of the Americas directly, instead of relying on information about the historical and demographic location from which the families have immigrated to infer their extent of experience with Indigenous ways. This would require more in-depth contact with the children's families and communities than was possible in the present school-based study.

It would also be valuable to further examine the generality of keen attention to surrounding events to examine its prevalence among Indigenous people on other continents. Ethnographic studies report that learning through observation is valued in many other communities around the world without a long history of schooling (Harkness, 1977; Mitchell-Kernan & Kernan, 1975; Ochs & Schieffelin, 1984; Ward, 1971). Systematic comparative studies focused on third-party attention in other cultural settings around the world would allow examination of the generality of the patterns, and their relation to communities' histories and people's participation in their communities' traditions and institutions.

It would also be fruitful to examine the community organization and practices that encourage learning by being aware of ongoing activities. Learning by observing surrounding activities depends on being present when notable events occur. What else is involved in allowing or encouraging attention to one's surroundings? And to what extent do the approaches of the children and families vary across settings (e.g., home, community, and school) that differ in their organization and practices?

The Role of Schooling (and Related) Approaches

Our findings also support the idea that family experience with the forms of interaction and learning common in Western schooling may compete with ways of organizing learning valued in many Indigenous-heritage communities of the Americas. Schooling experience appears to be an important feature of a cluster of practices that may relate to how children approach learning (Rogoff & Angelillo, 2002). For example, greater maternal schooling has been associated with smaller family size and thus fewer opportunities to learn by observing others, as well as differences in occupation, language, and other cultural practices (LeVine, 1987; LeVine et al., 2001; Najafi et al., 2007; Rogoff et al., 2005).

Experience with Western school has been related to the ways that parents interact with their children in the home (Crago, Annahatak, & Ningiuruvik, 1993; Laosa, 1980, 1982; Richman, Miller, & LeVine, 1992; Rogoff, 2003). Like in school, where the teacher often "orchestrates" individual and group attention (O'Connor & Michaels, 1993, 1996), mothers with extensive experience with school are more likely to interact with children in school-like ways, such as managing children's attention and participation (Chavajay & Rogoff, 2002; Rogoff et al., 1993). Children from families with extensive schooling may rely more on having their attention directed by adults, and muster less attention when no one is telling them to attend, as in our activity.

In future research, it would be fruitful to examine whether in supportive circumstances, people may be able to fluently use multiple approaches to learning, or whether schooling experience necessarily or usually competes with ways of organizing learning in Indigenous-heritage communities of the Americas. Understanding the circumstances that promote expansion of people's repertoires (rather than replacement of one approach by another) would be important information for supporting teaching and learning for immigrant and other populations.

Conclusions and Implications

Our research supports the idea that children whose families have experience with practices of Indigenous-heritage communities of the Americas may be especially keen observers of events occurring around them. Such keen observation may be an important way to approach learning in communities where children are integrated in the range of activities of the community and pitch in when they are ready, in the tradition of learning through *intent community participation* (Rogoff et al., 2003; Rogoff et al., 2007).

The study underlines the importance of focusing on children's community practices as a way of understanding cultural variation, and not just relying on nationality or ethnicity as homogenous categories of culture. Had this study combined the children of the two backgrounds because they are both of Mexican descent, we would have missed important differences in approaches to learning. Rather than assuming that all Latinos or all Mexicans have the same characteristics, it is important for research to examine empirically the extent to which communities within such broad categories differ from as well as resemble each other. This is facilitated by consideration of cultural practices as ways of organizing life, rather than reducing cultural research to simply comparing national or ethnic categories.

Recognizing various ways that children approach learning activities can help in the organization of schools. Children come to school already familiar with practices in their communities that guide them in what is important to learn and how to learn. In addition, teachers come with their own values and methods of teaching influenced by their communities and their own history of involvement in Western school.

Children from communities with Indigenous roots may have to integrate several methods of learning into their repertoires in order to succeed in school. This may be both a challenge and an opportunity for the children as well as for schools, because being able to learn in several ways is likely to be useful in later life and other circumstances. Schools could make use of a likely propensity for keen observation among children with familiarity with Indigenous practices, along with teaching them how to engage in the adult-directed attentional management common in schools.

More broadly, observing ongoing interactions in which one is not involved may be beneficial to all children as a method of learning and being involved in community life. Indeed, schools could make greater use of children's observation of others as a pedagogical tool that may benefit all children's learning.

References

Akhtar, N. (2005). The robustness of learning through overhearing. *Developmental Science*, *8*, 199–209.

Barton, M. E., & Tomasello, M. (1991). Joint attention and conversation in mother-infant sibling triads. *Child Development*, *62*, 517–529.

Bonfil Batalla, G. (1988). Panorama étnico y cultural de México. In R. Stavenhagen (Ed.), *Política cultural para un país multiétnico* (pp. 61–68). México, DF: Secretaría de Educación Pública.

Bonfil Batalla, G. (1996). *México profundo: Reclaiming a civilization*. Austin, TX: University of Austin.

Cancian, F. M. (1964). Interaction patterns in Zinacanteco families. *American Sociological Review*, *29*, 540–550.

Chamoux, M.-N. (1992). Aprendiendo de otro modo [Learning in another way]. In M.-N. Chamoux (Ed.), *Trabajo, técnicas y aprendizaje en el México indígena* (pp. 73–93). Mexico City, Mexico: Centro de Investigaciones y Estudios Superiores en Antropología Social, Ediciones de la Casa Chata.

Chavajay, P. (1993). Afterward: Independent analysis of cultural variations and similarities in San Pedro and Salt Lake. In B. Rogoff, J. Mistry, A. Goncu, & C. Mosier (Eds.), Guided participation in cultural activity by toddlers and caregivers. *Monographs of the Society for Research in Child Development*, *58*(Serial No. 236), 162–165.

Chavajay, P. (2006). How Mayan mothers with different amounts of schooling organize a problem-solving discussion with children. *International Journal of Behavioral Development*, *30*, 371–382.

Chavajay, P., & Rogoff, B. (1999). Cultural variation in management of attention by children and their caregivers. *Developmental Psychology*, *35*, 1079–1090.

Chavajay, P., & Rogoff, B. (2002). Schooling and traditional collaborative social organization of problem solving by Mayan mothers and children. *Developmental Psychology*, *38*, 55–66.

Corona, Y., & Pérez, C. (2005, June). *The enrichment of community relations through children's participation in ceremonial life*. Paper presented at Childhoods 2005:

Children and Youth in Emerging and Transforming Societies, Oslo, Norway.

Correa-Chávez, M., & Rogoff, B. (2009). Children's attention to interactions directed to others: Guatemalan Mayan and European American patterns. *Developmental Psychology, 45*, 630–641.

Correa-Chávez, M., Rogoff, B., & Mejía Arauz, R. (2005). Cultural patterns in attending to two events at once. *Child Development, 76*, 664–678.

Crago, M. B., Annahatak, B., & Ningiuruvik, L. (1993). Changing patterns of language socialization in Inuit homes. *Anthropology and Education Quarterly, 24*, 205–223.

de Haan, M. (1999). *Learning as cultural practice: How children learn in a Mexican Mazahua community*. Amsterdam: Thela Thesis.

de Leon, L. (2000). The emergent participant: Interactive patterns in the socialization of Tzotzil (Mayan) infants. *Journal of Linguistic Anthropology, 8*, 131–161.

Dixon, S. D., LeVine, R. A., Richman, A., & Brazelton, T. B. (1984). Mother-child interaction around a teaching task: An African-American comparison. *Child Development, 55*, 1252–1264.

Dunn, R., Griggs, S., & Price, G. (1993). Learning styles of Mexican-American and Anglo-American elementary school students. *Journal of Multicultural Counseling and Development, 21*, 237–247.

Ellis, S., & Gauvain, M. (1992). Social and cultural influences on children's collaborative interactions. In L. T. Wineger & J. Valsiner (Eds.), *Children's development within social context* (Vol. 2, pp. 155–180). Hillsdale, NJ: Erlbaum.

Frye, D. (1996). *Indians into Mexicans: History and identity in a Mexican town*. Austin: University of Texas Press.

Gaskins, S. (2000). Children's daily activities in a Mayan village: A culturally grounded description. *Cross-Cultural Research, 34*(4), 375–389.

Greenfield, P. (2004). *Weaving generations together: Evolving creativity among the Mayas of Chiapas*. Santa Fe, NM: School of American Research Press.

Gutiérrez, K., & Rogoff, B. (2003). Cultural ways of learning: Individual traits or repertoires of practice. *Educational Researcher, 32*, 19–25.

Harkness, S. (1977). Aspects of social environment and first language acquisition in rural Africa. In C. Ferguson & C. Snow (Eds.), *Talking to children, Language input and acquisition* (pp. 309–316). New York: Cambridge University Press.

Heath, S. B. (1983). *Ways with words*. New York: Cambridge University Press.

Laosa, L. (1980). Maternal teaching strategies in Chicano and Anglo-American families: The influence of culture and education on maternal behavior. *Child Development, 51*, 759–765.

Laosa, L. (1982). School, occupation, culture, and family: The impact of parental schooling on the parent-child relationship. *Journal of Educational Psychology, 74*, 791–827.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.

LeVine, R. (1987). Women's schooling, patterns of fertility and child survival. *Educational Researcher, 16*, 21–27.

LeVine, R. A., LeVine, S. E., & Schnell, B. (2001). "Improve the women": Mass schooling, female literacy, and worldwide social change. *Harvard Educational Review, 71*, 1–35.

LeVine, R., & White, M. (1992). The social transformation of childhood. In A. S. Skolnick & J. H. Skolnick (Eds.), *Family in transition: Rethinking marriage, sexuality, child rearing, and family organization* (7th ed., pp. 295–315). New York: HarperCollins.

Lewis, M., & Feiring, C. (1981). Direct and indirect interactions in social relationships. In L. P. Lipsett (Ed.), *Advances in infancy research* (Vol. 1, pp. 129–161). Norwood, NJ: Ablex.

Lorente y Fernández, D. (2006). Infancia nahua y transmisión de la cosmovisión: Los *ahuaques* o espíritus pluviales en la Sierra de Texcoco, México [Nahua childhood and transmission of world view: The *ahuaques* or rain spirits in the Sierra of Texcoco, México]. *Boletín de Antropología Universidad de Antioquia, 20*, 152–168.

Martini, M. (1996). "What's new?" at the dinner table: Family dynamics during mealtimes in two cultural groups in Hawaii. *Early Development and Parenting, 5*, 23–34.

Mejía Arauz, R., Rogoff, B., & Paradise, R. (2005). Cultural variation in children's observation during a demonstration. *International Journal of Behavioral Development, 29*, 282–291.

Meyer, J., Ramirez, F., & Soysal, Y. (1992). World expansion of mass education, 1870–1980. *Sociology of Education, 65*, 128–149.

Mitchell-Kernan, C., & Kernan, K. T. (1975). Children's insults: America and Samoa. In M. Sanchez & B. Blount (Eds.), *Sociocultural dimensions of language use* (pp. 307–315). New York: Academic Press.

Modiano, N. (1973). *Indian education in the Chiapas highlands*. New York: Holt, Rinehart & Winston.

Morelli, G. A., Rogoff, B., & Angelillo, C. (2003). Cultural variation in young children's access to work or involvement in specialized child-focused activities. *International Journal of Behavioral Development, 27*, 264–274.

Moreno, R. P. (2000). Teaching practices of Mexican American mothers with everyday and school related tasks. *Merrill-Palmer Quarterly, 46*, 613–631.

Najafi, B., Mejía Arauz, R., & Rogoff, B. (2007, April). *Cultural continuities and transformations in Western and Indigenous Mexican practices*. Chicago: American Educational Research Association.

Ochs, E. (1988). *Culture and language development: Language acquisition and language socialization in a Samoan village*. Cambridge, UK: Cambridge University Press.

Ochs, E., & Schieffelin, B. B. (1984). Language acquisition and socialization: Three developmental stories and their implications. In R. Schweder & R. LeVine (Eds.), *Culture and its acquisition* (pp. 276–320). Chicago: University of Chicago Press.

O'Connor, M. C., & Michaels, S. (1993). Aligning academic task and participation status through revoicing: Analysis of a classroom discourse strategy. *Anthropology and Education Quarterly, 24*, 318–335.

O'Connor, M. C., & Michaels, S. (1996). Shifting participant frameworks: Orchestrating thinking practices in group discussions. In D. Hicks (Ed.), *Discourse, learning, and schooling* (pp. 63–103). New York: Cambridge University Press.

Oshima-Takane, Y., Goodz, E., & Deverensky, J. L. (1996). Birth-order effects on early language development: Do second born children learn from overheard speech? *Child Development, 67*, 621–634.

Paradise, R. (1996). Passivity or tacit collaboration: Mazahua interaction in cultural context. *Learning and Instruction, 6*, 379–389.

Philips, S. (1972). Participant structures and communicative competence: Warm Spring children in community and classroom. In C. B. Cazden, V. P. John & J. Hymes (Eds.), *Functions of language in the classroom*. New York: Teachers College Press.

Repacholi, B. M., & Meltzoff, A. N. (2007). Emotional eavesdropping: Infants selectively respond to indirect emotional signals. *Child Development, 78*, 503–521.

Richman, A. L., Miller, P. M., & LeVine, R. A. (1992). Cultural and educational variation in maternal responsiveness. *Developmental Psychology, 28*, 614–621.

Rogoff, B. (2003). *The cultural nature of development*. New York: Oxford University Press.

Rogoff, B., & Angelillo, C. (2002). Investigating the coordinated functioning of multifaceted cultural practices in human development. *Human Development, 45*, 211–225.

Rogoff, B., Correa-Chávez, M., & Navichoc Cotuc, M. (2005). A cultural/historical view of schooling in human development. In D. B. Pillemer & S. H. White (Eds.), *Developmental psychology and social change: Research, history, and policy* (pp. 225–263). New York: Cambridge University Press.

Rogoff, B., Mistry, J., Göncü, A., & Mosier, C. (1993). Guided participation in cultural activity by toddlers and caregivers. *Monographs of the Society for Research in Child Development, 58*(Serial No. 236).

Rogoff, B., Moore, L., Najafi, B., Dexter, A., Correa-Chávez, M., & Solís, J. (2007). Cultural routines and practices. In J. Grusec & P. Hastings (Eds.), *The handbook of socialization: Theory and research* (pp. 490–515). New York: Guilford.

Rogoff, B., Paradise, R., Mejía Arauz, R., Correa-Chávez, M., & Angelillo, C. (2003). Firsthand learning through intent participation. *Annual Review of Psychology, 54*, 175–203.

Schieffelin, B. B. (1991). *The give and take of everyday life: Language socialization of Kaluli children*. Cambridge, UK: Cambridge University Press.

Stavenhagen, R. (1988). Introducción. In R. Stavenhagen (Ed.). *Política cultural para un país multiétnico* (pp. 7–21). México, DF: Secretaría de Educación Pública.

Urrieta, L., Jr. (2003). *Las identidades también lloran*: Identities also cry: Exploring the human side of Indigenous Latina/o identities. *Educational Studies, 34*, 147–168.

Verhoef, H., Morelli, G., & Anderson, C. (1999). *"Please don't interrupt me, I'm talking": Cultural variation in toddlers' attention seeking efforts and caregivers' responses*. Manuscript submitted for publication.

Vigil, J. D. (1998). *From Indians to Chicanos: The dynamics of Mexican-American culture*. Prospect Heights, IL: Waveland Press.

Ward, M. C. (1971). *Them children: A study in language learning*. New York: Holt, Rinehart & Winston.

Whiting, B. B., & Edwards, C. P. (1988). *Children of different worlds: The formation of social behavior*. Cambridge, MA: Harvard University Press.

Zelizer, V. A. (1992). Pricing the priceless child: From baby farms to black-market babies. In A. S. Skolnick & J. H. Skolnick (Eds.), *Family in transition* (pp. 295–315). New York: HarperCollins.

Child Development, May/June 2010, Volume 81, Number 3, Pages 913–928

Children's Understanding of Social-Cognitive and Social-Communicative Aspects of Discourse Irony

Eva Filippova and Janet Wilde Astington
University of Toronto

To bridge the social-reasoning focus of developmental research on irony understanding and the pragmatic focus of research with adult populations, this cross-sectional study examines 5-, 7-, and 9-year-olds' ($n = 72$) developing understanding of both social-cognitive and social-communicative aspects of discourse irony, when compared with adults ($n = 24$). Although 5-year-olds lag behind the other age groups in their reasoning about the speaker's meaning, belief, intention, and motivation, adults are consistently superior to children of all ages on these social-cognitive measures. In contrast, limited age-related differences were found in participants' judgment of the social-communicative function of irony (how nice, mean, and funny irony is). The findings help to reconcile previous discrepant claims as to the age when children come to understand irony.

Discourse irony is a crafty strategy for the negotiation of meaning in social contexts and for carrying on social relationships. All forms of irony—whether sarcasm, humorous bantering, jocularity, understatement, overstatement, and so on—involve indirectly communicated beliefs and attitudes that constitute a speaker's implied meaning. For instance, an ironist can say "what a gracious dive" to comment on a friend's landing on the pool surface with her belly. The listener must substitute the intended meaning for the literal one uttered. In addition, an appreciation of irony entails the listener's recognition of the speaker's motivation for such meaning substitution. For instance, the use of "sarcasm" is motivated by a desire to ridicule someone while elevating one's own position whereas the use of "jocularity" is motivated by the speaker's wish to convey group solidarity (e.g., Gibbs, 2000).

Full appreciation of irony involves an appreciation of the social-cognitive as well as the social-communicative aspects of irony. Although the former refers to an assessment of the mind of the speaker, the latter refers to the pragmatic function of irony, that is, what the use of irony achieves for the speaker in a conversation. Within the field of children's theory of mind, researchers have used irony as a means of investigating the development of higher order social-cognitive abilities (e.g., Happé, 1993, 1995; Sullivan, Winner, & Hopfield, 1995; Winner & Leekam, 1991). In contrast, studies examining social-communicative aspects of irony have primarily examined adults' appreciation of irony's pragmatic function, without exploring what cognitively underlies this appreciation (e.g., Clark & Gerrig, 1984; Grice, 1975; Kreuz & Glucksberg, 1989; Kumon-Nakamura, Glucksberg, & Brown, 1995; Searle, 1979; Sperber & Wilson, 1995). However, a complete understanding of irony involves an appreciation of both social-cognitive (i.e., theory of mind) and social-communicative (i.e., pragmatic-functional) aspects. The present study examines both of these facets of irony interpretation within the same research paradigm. In so doing, our study unites the social-cognitive focus of the developmental research and the pragmatic focus of irony research in adults in a single developmental investigation in order to reveal developmental perspectives on both aspects of irony understanding.

Investigating particularities in the development of understanding of these different aspects of irony may help reconcile claims that 5- and 6-year-olds understand some aspects of irony (e.g., Dews et al., 1996, for irony muting function; Harris & Pexman,

---

This research was supported by doctoral fellowships from the Social Sciences and Humanities Research Council of Canada and the Ontario Graduate Scholarship Program to EF and a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada to JWA. We thank Sandra E. Trehub and Jens Brockmeier for their input throughout the study, Sue Elgie and Jan Volin for statistical advice, and Karen Milligan, Charlie Lewis, and three anonymous reviewers for their comments on earlier drafts of the manuscript.

Eva Filippova is now at The Institute of Phonetics, Charles University in Prague, Prague, Czech Republic.

Correspondence concerning this article should be addressed to Eva Filippova, Institute of Phonetics, Faculty of Arts, Charles University in Prague, Nám. J. Palacha 2, 116 38 Prague 1, Czech Republic. Electronic mail may be sent to eva.filippova@ff.cuni.cz.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8103-0018

2003; Sullivan et al., 1995; Winner & Leekam, 1991) and the counterclaim that even children in middle-school years do not understand irony (e.g., Demorest, Meyer, Phelps, Gardner, & Winner, 1984; Demorest, Silberstein, Gardner, & Winner, 1983, for conflation of irony and deception). Indeed, even young children may perceive the meanness of a sarcastic comment or appreciate the humor of a counterfactual assertion without explicitly figuring out that the speaker has said something he or she believes to be untrue and does not intend the listener to believe it to be true.

The Present Study

Children's understanding of both the social-cognitive and social-communicative aspects of irony was expected to improve through the early- to middle-school years. Tracing development of understanding of each of these aspects was the main goal of this study. Children heard stories in which a character made an ironic statement and were asked (a) to make judgments about the speaker's mental states (belief, intention, attitude) and (b) to evaluate the pragmatic function of the statement (how nice, mean, funny it was). Thus, we can reveal any differences in the development of understanding of the social-cognitive and social-communicative aspects of irony (a and b, respectively). We may show that before children explicitly conceptualize speakers' beliefs and intentions, they can perceive meanness and humor in nonliteral statements. We operationalize and refer to these statements as ironic, while acknowledging that in some cases children may not have conceptualized them as irony based on their judgment of speaker beliefs and intentions.

Children's comprehension of counterfactual assertion is documented to emerge around the age of 6 years (e.g., Ackerman, 1983; Andrews, Rosenblatt, Malkus, Gardner, & Winner, 1986; Dews et al., 1996; Harris & Pexman, 2003; Winner & Leekam, 1991; Winner et al., 1987). Although these studies systematically tested irony in the form of counterfactual statements (i.e., statements in which the implied meaning is the exact opposite of the expressed meaning), the task in the present study also incorporates irony in the form of noncounterfactual statements. As deliberately exaggerated statements (hyperbole) are also considered ironic (Gibbs, 2000; Leggitt & Gibbs, 2000; Winner et al., 1987), the present study uses hyperbole as a contrasting form to counterfactual statements. In other words, although counterfactual statements communicate the opposite valence to that uttered, hyperbolic statements communicate the same valence as that uttered. All the forms of irony used in this study—counterfactual praise or "jocularity," counterfactual criticism or "sarcasm," and hyperbole—are the most frequently occurring types of irony in everyday adult discourse (e.g., Gibbs, 2000). Additionally, all these types of utterances are less dependent on direct echoing of previously mentioned thoughts (i.e., explicitly referring to something said earlier; Gibbs, 2000). The present design complements Hancock, Dunham, and Purdy's (2000) study that revealed a bias to detect ironic criticism before praise, to the latter of which echoing contributes significantly. Our task omits the explicit echoing to reveal any underlying aspects of implicit reference to social norms or expectations. Thus, the design of the task enables a direct comparison of children's and adults' interpretation of critical and praise forms of irony occurring in two types of statements—counterfactual statements and noncounterfactual exaggerated statements (i.e., hyperbole).

Along with our primary developmental focus on understanding irony, our study examines whether forms of ironic criticism are understood before forms of ironic praise (Hancock et al., 2000). It also explores whether counterfactual assertions are understood before noncounterfactual assertions, a hypothesis based on the pronounced saliency of counterfactual statements (i.e., opposite of the logically expected statement) in contrast to the less salient noncounterfactual statements (Creusere, 2000; Demorest et al., 1983). Thus, an inclusion of critical *and* praise valence of irony in the form of counterfactual *and* noncounterfactual assertions enables an exploration of the documented asymmetry in understanding different kinds of discourse irony.

Further, different types of ironic statement serve different pragmatic functions and communicate different moods and attitudes. Our study also explores the understanding of social-communicative (i.e., pragmatic) functions of irony identified by children and adults. As hyperbolic criticism (HC) is a directly offensive type of statement, we expected it to be judged as the meanest statement of all. We expected counterfactual criticism (CC), due to its muting critical tone relative to exaggerated critical comments, to be judged as the second meanest type of statement, followed by counterfactual praise (CP), itself followed by hyperbolic praise (HP). The reverse order was hypothesized for the rating on the niceness scale. Lastly, as humor is one of the goals of irony (e.g., Creusere, 2000; Dews et al., 1996; Gibbs, 2000; Kumon-Nakamura et al., 1995; Pexman, Glenwright, Krol, & James, 2005), we expected the statements to be judged as funny, especially in the counterfactual types of statements where the context–utterance discrepancy and ambiguity are more salient than in hyperbole.

Method

Participants

Children of three different ages (*n* = 72) participated in the study: twenty-four 5-year-olds (11 boys and 13 girls; age range = 5.2–6.2, *M* age = 5.8), twenty-four 7-year-olds (13 boys and 11 girls; age range = 7.2–8.1, *M* age = 7.7), and twenty-four 9-year-olds (11 boys and 13 girls; age range = 9.1–10.1, *M* age = 9.7). All the children came from middle- to upper-middle-income areas of a major Canadian city and were at the same school. Children were recruited through their class teachers. Due to the nature of projected follow-up comparative research with different language groups, the present sample was ethnically and linguistically homogeneous, consisting of monolingual English speakers only.

In addition, 24 adults (11 men and 13 women; age range = 23–63, *M* age = 32) participated in the study. The adult sample consisted of 10 monolingual English speakers, 13 speakers of more than one language whose maternal tongue was English, and 1 speaker with English as a second language yet fully proficient in English. The adult participants were from a range of educational and professional backgrounds. They were recruited through friends and acquaintances of colleagues of the investigators.

Materials and Design

The task consisted of eight short stories, each describing an everyday interaction between a child and another person, and ending with an ironic comment made by the other person to the child. Each story was illustrated by one picture, in which the other person was depicted from behind, facing the child, to eliminate any mediation of interpretative cues through facial expressions (see the Appendix for an example of story and illustration).

Each of these eight stories had two possible kinds of ending—a positive and a negative—to accommodate the valence of the ironic comment in the form of ironic praise and ironic criticism, respectively. Further, the form of the ironic comment at the end of each story was of two possible types: either a counterfactual assertion or a non-counterfactual, hyperbolic assertion.

The two factors (i.e., valence and type) were crossed to yield four different versions of the eight unique scenario types (32 stories in total). Each participant's protocol consisted of one version of each scenario type, for a total of eight different stories (i.e., no protocol repeated the same story content with a different ending or different ironic comment). Each protocol included two CCs, two HCs, two CPs, and two HPs (see Table 1 and story example in the Appendix). All the statements of all eight stories were completely counterbalanced, as was their order across the testing protocols.

Following each irony scenario, there were two comprehension questions followed by two sets of test questions: the first set assessed social-cognitive understanding (i.e., the detection of the meaning, the speaker's first-order belief, third-order intention, and the speaker's motivation and/or attitude); the second set of questions examined participants' judgment of the social-communicative function of the statement (i.e., how nice, how mean, and how funny the statements were) on a scale from 0 (*not at all*) to 3 (*very very*; see the Appendix for an example of story and questions).

Procedure

Children were interviewed individually in two sessions, taking place on two different days in a quiet room on their school premises. The procedure was identical for all participating children. At the beginning of the first session, children were trained on how to use the rating scale, depicted as four glasses with an increasing amount of liquid (*not full at all*, for not nice/mean/funny at all, to *very very full*, for very very nice/mean/funny).

Children were presented with four stories per session, with one CC, one CP, one HC, and one HP in each session. In addition to the task reported here, children also received advanced theory-of-mind, language, memory, and prosody tasks, as

Table 1
Conditions of the Irony Task

	Valence	
Type	Criticism	Praise
Counterfactual	CC	CP
Hyperbolic	HC	HP

Note. CC = counterfactual criticism; CP = counterfactual praise; HC = hyperbolic criticisms; HP = hyperbolic praise.

reported in Filippova and Astington (2008). Thus, another type of task was administered between each irony story. Each session lasted for up to 40 min.

During the administration of the task, the child listened to the experimenter reading the individual stories and looked at the accompanying illustrations. The experimenter used consistently the same stressed intonation for the ironic utterances across all participating children and made no facial expressions. This resulted in all children receiving the same prosodic cues and no visual cues. Importantly, this personalized way of administration of tasks to children, common in the experimental tradition of theory-of-mind research, not only lessens the information processing load imposed on a child to abstract the facts out of a recorded story but also offers a direct and personal way of approaching each child. Furthermore, whenever a parent's consent for video-recording the session was given, the consistency of prosody and visual cues was later checked from the videotape: The researchers and the transcriber checked for any deviations from the standard prescribed prosody patterns and for no eye contact or mimics while the ironic statements were read. No inconsistencies were found.

Passing both comprehension questions was a critical prerequisite for inclusion in the data analysis; if a child failed to respond correctly to either of the control questions, the story was reread and the two comprehension questions were asked once more to ensure that children understood the key story facts and remembered the ironic comment. None of the participants failed any control questions twice and so all participants and all trials are included in the data analysis. Subsequently, a set of four social-cognitive questions (i.e., detection of meaning, belief, communicative intention, and motivation of the speaker) was asked, followed by a set of three social-communicative questions (i.e., how nice, mean, and funny the statement was). Children were allowed to take as many breaks throughout the session as they needed, having also an opportunity to select their prizes (i.e., stickers) as their reward and motivation for further testing. Only the youngest participants required a short break or two.

The same eight irony stories were used across the adult volunteers, each of whom was asked to fill out eight questionnaires according to the short stories that preceded them. The stress in the ironic statements was marked and the protocols were identical to those used with the child participants (see the Appendix). In the instructions preceding the eight questionnaires, the adults were asked to read the statements out loud with a stress placed on the CAPITALIZED WORDS and to take breaks if they happened to get tired or bored by the format of the questions. As with the children, the order of the statements and stories was completely counterbalanced. Although the adult group is not, strictly speaking, a control group, it provides a sufficient approximation of the trend of adult answers to the questions of the study and a reasonable comparison between child and adult data.

Scoring and Coding

Individual scores for the social-cognitive aspects of irony understanding were obtained from each participant's answers to the meaning, belief, intention, and motivation questions in each tested scenario. The dichotomous yes–no questions investigating the meaning, belief, and intention of the ironist were scored as either *correct* (1) or *incorrect* (0). Based on our previous research (Filippova & Astington, 2008), participants' answers to the motivation question of each story were coded on a 5-point scale, reflecting levels of interpersonal understanding based on increasing cognitive complexity of the answers, identified from the patterns of adults' responses: a score of 0 was assigned for answers irrelevant to the identification of the speaker motivation (e.g., *he said it by accident*), for wrong answers, or two consecutive *I don't know* answers; a score of 1 was assigned for simple surface-level justifications, such as reference to relevant story facts (e.g., *because he spilled the water; because she watered her flowers*), to feelings of the speaker (e.g., *because she was mad at him; because she was happy about what he has done*), or for learned conventional answers (e.g., *that's not what you do at a store, that was a nice thing to do*); a score of 2 was assigned for a reference made to information that had an implication for another person's behavior (e.g., *she didn't want him to do it again*); a score of 3 was assigned for identification of the speaker's intention to affect another person's mental state rather than behavior (e.g., *she wanted to make him feel better about what he did*); and a score of 4 was assigned for a reference made to a speaker's attitude toward a person or a situation (e.g., *he was joking; she was mocking him*) or to the pragmatic function of irony that goes beyond the identification of the speaker's intention (e.g., *he was lightening the situation; he was adding humor; he was being sarcastic*). (This generalized coding is derived from 32 coding subschemes developed for each version of each story, based on the patterns of adult responses to the motivation question;

therefore, an answer "he was mad" would be coded as 1 if given in the critical scenarios but would be coded as 0 if given in the praise scenarios.) Based on the scoring procedures used in theory-of-mind literature, whenever a participant gave an answer reflecting more than a single level of the scale, he or she was credited for their best answer to this question by being given the highest possible score on the motivation question.

Finally, individual scores for the social-communicative aspects of irony understanding were obtained from each participant's rating of the ironic statements on the niceness, meanness, and funniness scales. A rated score (0–3) was recorded from each participant, on each of the stories, and on each of the three scales.

The responses of every child participant on each story were scored and coded by two independent scorers: one of the investigators and an assistant, naïve to the hypotheses being tested and blind to the ages of the participants. Twenty percent of the responses of adult participants were checked by an independent coder against the scores and coding of the full set of adult data coded by one of the investigators. Interrater reliability of the coded answers ranged from 92% to 100% agreement for the individual scenarios (Cohen's kappas ranged from .82 to 1). All discrepancies were resolved to a final resolution accepted by both coders.

Results

Analysis of variance (ANOVA), examining the effects of different factors on participants' judgment of social-cognitive and social-communicative variables showed that neither sex (2) nor order (12) yielded a significant main effect or an interaction. Thus, data from boys and girls in all orders were combined in the following analyses.

There was some concern that the information-processing load of the task was high, particularly for 5-year-olds. In order to allay this concern, an analysis was conducted to compare 5-year-olds' performance on the set of questions across the first two stories and the second two stories that each child received in each session. Repeated measures analyses on the individual variables were computed with valence (criticism/praise) and type of statement (counterfactual/hyperbolic) as within-subject factors and sequence (first two stories received in the sessions/second two stories received in the sessions) as between-subject factors. There were no significant main effects of sequence (first or second two stories of the given sessions) and no significant interactions of sequence and valence (criticism vs. praise) or type of statement (counterfactual vs. hyperbolic). Thus, the 5-year-olds' performance is comparable across the sequence of stories they heard first and later in the session. These findings demonstrate that an information-processing challenge was not occurring as the duration of the session increased and more information was presented. Thus, data from 5-year-olds are included in all analyses reported.

Repeated measures analyses determined the effect of valence (criticism vs. praise) and type of statement (counterfactual vs. hyperbole) as the within-subject factors and the effect of age (5-, 7-, 9-year-olds, and adults) as the between-subject factor.

Social-Cognitive Variables

As scores on three of four social-cognitive variables (i.e., meaning, belief, and intention) were dichotomous, an aggregate score was required for subsequent repeated-measures analyses. Reliability of individual measures was considered over the four different conditions (see Table 1): Valence (criticism vs. praise) × Type (counterfactual vs. hyperbole) with the resulting alphas: meaning, $\alpha = .80$; belief, $\alpha = .79$; intention, $\alpha = .79$. Because of an acceptable reliability on all measures, aggregate scores were generated from the dichotomous scores on the meaning, belief, and intention questions by summing the performance scores of each child across the two versions of the same condition. Although scores on the motivation variable ranged from 0 to 4, for the sake of consistency, it was treated in the same way. Reliability of the motivation measure over the four different conditions was $\alpha = .83$.

To examine the hypothesized age-related improvement on the irony task as well as the effect of valence of ironic statements—criticism versus praise—and type of statements—counterfactual versus hyperbole—a three-way repeated measures ANOVA was conducted on the meaning, belief, intention, and motivation variables across four conditions resulting from the factors of valence and type being crossed: CC, CP, HC, and HP.

In the repeated measures analysis, valence and type were crossed within-subject factors. The valence factor had two levels (criticism and praise), as did the type factor (counterfactual and hyperbole). Age (5-, 7-, 9-year-olds, and adults) was a between-subject factor. The Valence × Age and Type × Age interaction effects were tested using

Table 2

Means and Standard Deviations for Social-Cognitive Variables for Four Story Conditions for Four Age Groups

		Praise				Criticism			
		Counterfactual		Hyperbolic		Counterfactual		Hyperbolic	
Age	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Meaning[a]									
5	24	1.29	0.86	1.04	0.91	1.17	0.82	0.79	0.72
7	24	1.75	0.44	1.58	0.72	1.75	0.53	1.71	0.55
9	24	1.50	0.66	1.83	0.56	1.87	0.45	1.75	0.61
Adults	24	1.87	0.34	2.00	0.00	1.96	0.20	1.96	0.20
Belief[a]									
5	24	1.17	0.92	0.58	0.83	1.17	0.82	0.92	0.83
7	24	1.87	0.34	1.29	0.81	1.79	0.42	1.29	0.62
9	24	1.75	0.61	1.21	0.88	1.83	0.48	1.29	0.69
Adults	24	1.83	0.38	1.92	0.28	1.92	0.28	1.83	0.48
Intention[a]									
5	24	1.42	0.83	0.54	0.72	0.58	0.65	1.17	0.76
7	24	1.71	0.62	0.63	0.82	1.17	0.70	1.08	0.78
9	24	1.46	0.88	0.71	0.91	0.88	0.80	1.00	0.83
Adults	24	1.79	0.51	1.33	0.70	1.83	0.48	1.75	0.61
Motivation[b]									
5	24	2.21	1.79	2.83	1.63	3.42	1.98	2.12	1.08
7	24	3.96	2.63	3.00	1.59	4.62	2.22	2.63	0.97
9	24	5.12	2.71	4.04	1.83	6.25	1.75	3.21	1.84
Adults	24	7.08	1.98	5.17	1.95	5.67	2.16	4.79	2.52

[a] Out of 2. [b] Out of 8.

the multivariate criterion of Wilks's lambda (Λ). The valence main effect, the type main effect, and the Valence × Type interaction effect were tested using univariate tests. Table 2 summarizes the mean scores and standard deviations for the individual variables.

Meaning. The ANOVA on the meaning variable revealed a significant effect of age, $F(3, 92) = 21.51$, $p < .001$, partial $\eta^2 = .41$. Dunnett *C* post hoc test indicated significant differences between 5-year-olds and each other group and between 7-year-olds and adults. The univariate tests associated with valence main effect, type main effect, and the Valence × Type interaction were nonsignificant. Multivariate tests revealed marginally significant interactions between valence and age, Wilks's $\Lambda = .92$, $F(3, 92) = 2.61$, $p = .056$, multivariate $\eta^2 = .08$, and between type and age, Wilks's $\Lambda = .92$, $F(3, 92) = 2.43$, $p = .07$, multivariate $\eta^2 = .07$. Pairwise simple-effect analyses revealed that 5-year-olds were significantly poorer than the other groups on valence and type ($p \leq .001$) and that the 7-year-olds were poorer than adults on hyperbole ($p < .05$) when judging the literalness of the meaning (see Figure 1a and b), a finding consistent with the literature. Thus, age was the only significant factor in deciphering the nonliteral nature of the ironic speaker's meaning.

Belief. The ANOVA on the variable belief indicated a significant effect of age, $F(3, 92) = 18.13$, $p < .001$, partial $\eta^2 = .37$. Dunnett *C* post hoc test revealed significant differences between 5-year-olds and all other groups and between adults and all child groups. Univariate tests indicated a significant main effect of type, favoring counterfactual statements, Wilks's $\Lambda = .73$, $F(1, 92) = 34.74$, $p < .001$, partial $\eta^2 = .27$, and nonsignificant main effects of valence and the Valence × Type interaction. Multivariate tests revealed a nonsignificant Valence × Age interaction but indicated a significant Type × Age interaction, Wilks's $\Lambda = .88$, $F(3, 92) = 40.75$, $p < .01$, multivariate $\eta^2 = .12$. Pairwise simple-effect comparisons of the Type × Age interaction showed significant differences between 5-year-olds and all other age groups on counterfactual statements (all at $p < .001$). In the case of hyperbole, adults differed from all other groups (all at $p \leq .001$) and 5-year-olds differed from both 7-year-olds and 9-year-olds ($p < .01$; see Figure 1c and d).

Thus, participants detected the ironic speaker's belief in counterfactual statements more readily than in hyperbole. Again, performance of 5-year-

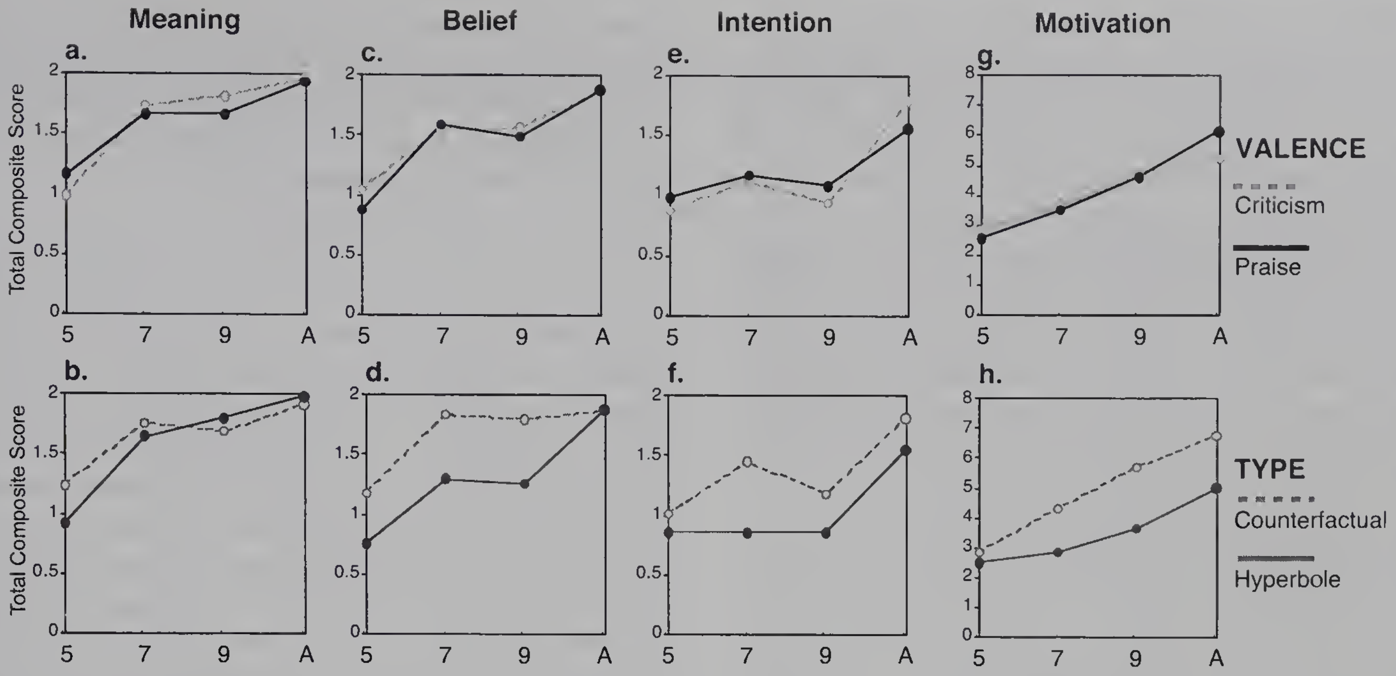


Figure 1. Performance on the social-cognitive variables for valence and type factors for the four age groups.
Note. Key for the x-axis: 5 = 5-year-olds; 7 = 7-year-olds; 9 = 9-year-olds; A = adults.

olds lagged behind that of the older groups. Although detection of the speaker's belief by 7- and 9-year-olds approached that of adults on counterfactual statements, their skills were poorer than those of adults on hyperbole.

Intention. The ANOVA on the intention variable revealed a significant effect of age, $F(3, 92) = 9.77$, $p < .001$, partial $\eta^2 = .24$. Dunnett *C* post hoc test showed significant differences between adults and all other age groups. Univariate tests indicated a nonsignificant main effect of valence. However, they revealed a significant type main effect, again favoring counterfactual statements, Wilks's $\Lambda = .77$, $F(1, 92) = 26.81$, $p < .001$, partial $\eta^2 = .23$, and a significant interaction between valence and type, Wilks's $\Lambda = .61$, $F(1, 92) = 59.21$, $p < .001$, partial $\eta^2 = .39$. Specifically, HP was the easiest condition for deciphering the intention of the speaker, followed by CC, and then HC, with CP being the most difficult condition of the four. Multivariate tests associated with the Valence × Age interaction and Type × Age interaction were nonsignificant, yet there was a significant three-way interaction between valence, type, and age, Wilks's $\Lambda = .90$, $F(3, 92) = 3.42$, $p < .05$. More precisely, adults were superior to all child groups in the cases of CC ($p \leq .001$), HC ($p < .01$), and HP ($p < .01$). In addition, 7-year-olds were significantly better than 5-year-olds in their understanding of the intention of a speaker using CC ($p < .01$). No significant differences among the groups were found in participants' recognition of the speaker's intention in HC, the condition in which the intent to criticize is the most salient of all the conditions used (see Figure 1e and f).

The detection of the speaker's intention occurred more readily, again, in counterfactual statements. Although 7-year-olds' performance exceeded that of 5-year-olds on counterfactual statements, even 9-year-olds performed no better than younger children on hyperbole. Adults' ability to discern the speaker's intention in both types of statements was superior to that of all age groups of children.

Motivation. The ANOVA on the motivation variable revealed a significant effect of age, $F(3, 92) = 20.78$, $p < .001$, partial $\eta^2 = .40$. Dunnett *C* post hoc test showed significant differences between adults and 5- and 7-year-olds and between 5- and 9-year-olds. Again, univariate tests showed a significant main effect of type, favoring counterfactual statements, Wilks's $\Lambda = .57$, $F(1, 92) = 68.17$, $p < .001$, partial $\eta^2 = .43$, but no significant main effect of valence. Furthermore, there was a significant interaction between valence and type, Wilks's $\Lambda = .91$, $F(1, 92) = 8.70$, $p < .01$, partial $\eta^2 = .09$, with the following order of increasing difficulty: CC, CP, HP, and HC. Multivariate tests indicated a significant Valence × Age interaction, Wilks's $\Lambda = .92$, $F(3, 92) = 2.86$, $p < .05$, multivariate $\eta^2 = .09$, and a significant interaction between type and age, Wilks's $\Lambda = .86$, $F(3, 92) = 5.09$, $p < .01$, multivariate $\eta^2 = .14$. In the case of counterfactual statements, pairwise simple-effect comparisons revealed significant differences between each group and every other group: 5- versus 7-year-olds ($p < .01$), 5- versus 9-year-olds and adults

($p < .001$), 7- versus 9-year-olds ($p < .01$), 7-year-olds versus adults ($p < .001$), and 9-year-olds versus adults ($p < .05$). In the case of hyperbole, pairwise simple-effect comparisons indicated significant differences between adults and all other groups ($p < .001$) and between 5- and 9-year-olds ($p < .01$).

Moreover, multivariate tests revealed a significant three-way interaction among valence, type, and age, Wilks's $\Lambda = .87$, $F(3, 92) = 4.55$, $p < .01$, multivariate $\eta^2 = .13$. The pairwise simple-effect comparisons revealed significant differences among age groups across the conditions: Adults differed from 5- and 7-year-olds on HP ($p < .001$), CP ($p < .001$), and HC ($p < .001$). The adults also differed from the 9-year-olds on HP ($p < .05$), CP ($p < .01$), and HC ($p < .01$). As well, 5-year-olds differed from 7-year-olds on CC ($p < .05$) and CP ($p < .05$), and from 9-year-olds on CC ($p < .001$), HC ($p < .05$), CP ($p < .001$), and HP ($p < .05$). Additionally, the 7- and 9-year-olds differed on CC ($p < .01$) and HP ($p < .05$; see Figure 1g and h).

Thus, participants were more sophisticated in judging an ironist's motivation in counterfactual statements than in hyperbole. More precisely, judging the motivation behind CCs by alluding to the speaker's mental states turned out to be easier than doing so for CP, which in turn, was easier than HP. HC represents the most difficult context for interpretation of the speaker's motivation for the use of irony. Nevertheless, there were pronounced differences across age in the ability to allude to the mental states of the speaker and his or her attitude toward the addressee or situation when asked to judge the ironist's motivation. There were clear indications of age-related changes in giving more elaborate references to speaker's higher order mental states with respect to an ironist's motivation between 5 years of age and adulthood.

To summarize, age was the most significant factor in the participants' judgment of the ironist's mind; whereas the 5-year-olds lagged behind the other groups overall, the adults were consistently superior to children of all ages, especially on the intention and motivation questions. This finding is consistent with a body of literature in that understanding the speaker's communicative intention and/or attitude continues to develop during the early school years. The adult scores are considerably higher than the 9-year-olds', suggesting that progress in understanding the social-cognitive aspects of irony is one that continues well into adulthood. Although counterfactual statements tended to offer an easier context for interpreting the mind of an ironist, hyperbole did so to a lesser extent. The valence of irony (i.e., criticism or praise) made significantly less difference in providing a context more readily interpretable for participants of all ages than the type of ironic statement. The age factor was persistently significant.

Social-Communicative (Pragmatic-Function) Variables

To examine the effects of the valence of irony (praise vs. criticism) and the type of ironic statement (counterfactual vs. hyperbole) on the participants' judgment of the pragmatic functions of the statements, a three-way repeated measures ANOVA was conducted on the niceness, meanness, and funniness variables across the four crossed conditions—CC, CP, HC, and HP. Aggregate scores, summed for each child across the two versions of the same condition ranging from 0 to 6, were used in the analysis.

In the following repeated measures analyses, valence and type were within-subject factors and age was a between-subject factor. Valence and type factors had two levels each (i.e., praise/criticism and counterfactual/hyperbole, respectively) and age had four levels (5-, 7-, 9-year-olds, and adults). The Valence × Age and Type × Age interaction effects were tested using the multivariate criterion of Wilks's lambda (Λ). The valence and type main effects, and the Valence × Type interaction were tested using univariate tests. Table 3 summarizes the mean scores and standard deviations for the three social-communicative variables.

Niceness. In the participants' judgment of the niceness of the ironic statements, the overall effect of age was not significant, $F(3, 92) = .46$, $p = .72$, partial $\eta^2 = .02$. Univariate tests indicated a significant main effect of valence, Wilks's $\Lambda = .31$, $F(1, 92) = 208.60$, $p < .001$, partial $\eta^2 = .69$, and a significant main effect of type, Wilks's $\Lambda = .66$, $F(1, 92) = 46.78$, $p < .001$, partial $\eta^2 = .34$. Praise was judged as significantly nicer than criticism and hyperbole as significantly nicer than counterfactual statements. There was also a significant interaction between valence and type, Wilks's $\Lambda = .34$, $F(1, 92) = 181.80$, $p < .001$, partial $\eta^2 = .64$: HP was judged as significantly nicer than both kinds of counterfactual statements and HC as the least nice statement of all.

In addition, multivariate tests revealed a significant interaction between valence and age, Wilks's $\Lambda = .70$, $F(3, 92) = 13.12$, $p < .001$, multivariate $\eta^2 = .30$. The pairwise simple-effect comparisons revealed the only difference between 5-year-olds and adults ($p < .001$) in the critical scenarios; in the praise

Table 3

Means and Standard Deviations for Social-Communicative Variables for Four Story Conditions for Four Age Groups

		Praise				Criticism			
		Counterfactual		Hyperbolic		Counterfactual		Hyperbolic	
Age	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Niceness									
5	24	0.79	1.22	4.88	1.83	3.38	2.30	0.50	1.22
7	24	2.08	1.95	5.50	0.83	1.96	2.14	0.79	1.14
9	24	2.13	1.94	5.46	0.88	2.04	2.14	0.79	1.22
Adults	24	3.75	1.62	5.33	0.82	0.79	1.02	0.67	1.05
Meanness									
5	24	4.92	1.56	1.21	2.17	2.58	2.30	4.83	1.69
7	24	3.21	1.84	0.25	0.74	3.33	2.30	4.71	1.33
9	24	3.33	2.16	0.13	0.34	3.79	1.82	4.96	1.20
Adults	24	1.04	1.23	0.04	0.20	3.71	1.65	3.17	1.31
Funniness									
5	24	1.08	1.91	2.13	1.96	1.25	1.82	1.17	1.79
7	24	1.50	1.41	2.54	1.59	0.92	1.41	0.79	1.25
9	24	1.38	1.74	2.00	1.50	1.08	1.41	1.50	1.82
Adults	24	2.54	1.28	2.71	1.54	1.33	1.34	1.83	1.27

Note. All scores out of 6.

scenarios, the 5-year-olds differed from 7- and 9-year-olds ($p < .01$) and from adults ($p < .001$) and the adults differed also from 7- and 9-year-olds ($p < .05$). although the interaction between type and age was not significant, the tests showed a significant three-way interaction among valence, type, and age, Wilks's $\Lambda = .74$, $F(3, 92) = 10.55$, $p < .001$, multivariate $\eta^2 = .26$. The pairwise simple-effect comparisons showed no significant differences between the groups in their judgment of the niceness in HC or HP conditions. In the CC condition, however, 5-year-olds differed from 7- and 9-year-olds ($p < .05$) and from adults ($p < .001$) and the adults also differed from 7- and 9-year-olds ($p < .05$). Similarly, in the CP condition, 5-year-olds differed from both the 7- and 9-year-olds ($p < .01$) and the adults differed from all the child groups ($p \leq .001$; see Figure 2a and b).

Confirming our prediction, the participants judged HC as the least nice of all statements and HP as the nicest, with both counterfactual statements being judged in between. Although even the youngest group recognized the lack of niceness in the exaggerated criticism expressed through hyperbole, they judged CCs as significantly nicer than did the other groups. The 5-year-olds also found HP less nice than the rest of the groups. The niceness of CP differed most significantly across the ages: Although adults judged such statements significantly nicer than the rest of the groups, the youngest group found such statements significantly less nice.

The lack of an overall pronounced difference among the age groups indicates that children, with the exception of the youngest age group in certain conditions, are fairly good at detecting how nice the ironic statements are, relative to adults. With the exception of HP, however, none of the age groups including the adults judged irony to be very nice. This suggests that irony in the forms used does not communicate overly positive attitudes to our group of respondents.

Meanness. The analysis of the meanness variable revealed a significant age effect, $F(3, 92) = 9.56$, $p < .001$, partial $\eta^2 = .24$. Dunnett *C* post hoc test revealed significant differences between adults and each child group. Univariate tests showed a significant main effect of valence, Wilks's $\Lambda = .36$, $F(1, 92) = 162.00$, $p < .001$, partial $\eta^2 = .64$, and a significant main effect of type, Wilks's $\Lambda = .71$, $F(1, 92) = 38.29$, $p < .001$, partial $\eta^2 = .29$. Criticism was judged as significantly meaner than praise and counterfactual statements as significantly meaner than hyperbole. There was also a significant interaction between valence and type, Wilks's $\Lambda = .40$, $F(1, 92) = 137.60$, $p < .001$, partial $\eta^2 = .60$. HC was judged as the meanest of all, followed by both kinds of counterfactual statements. HP was judged as the least mean of all statements.

Multivariate tests revealed a significant interaction between valence and age, Wilks's $\Lambda = .79$, $F(3, 92) = 9.26$, $p < .001$, multivariate $\eta^2 = .23$. The

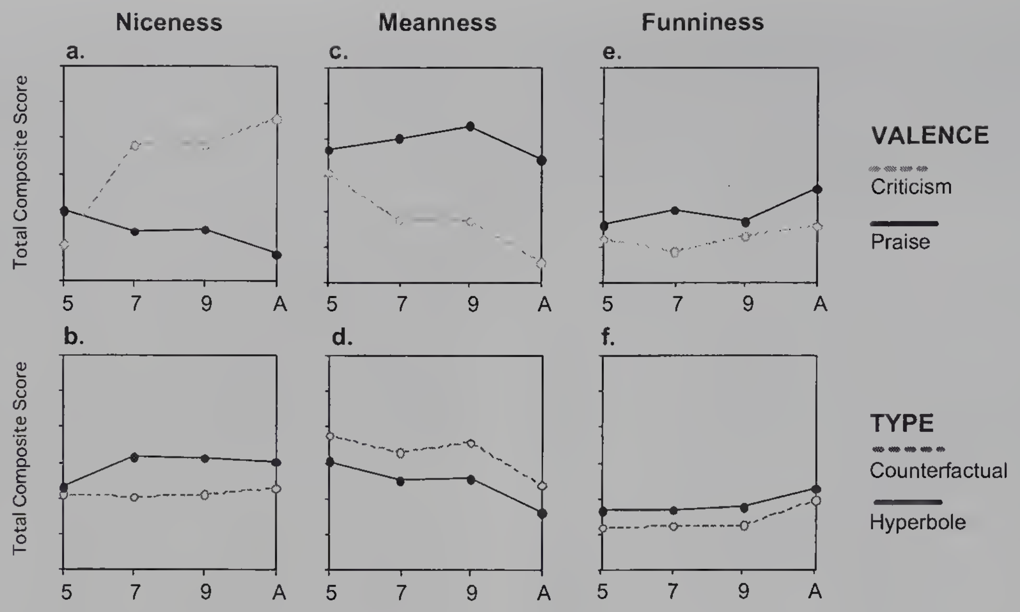


Figure 2. Performance on the social-communicative variables for valence and type factors for the four age groups.
Note. Key for the *x*-axis: 5 = 5-year-olds; 7 = 7-year-olds; 9 = 9-year-olds; A = adults.

pairwise simple-effect comparisons revealed that only the difference between the adults and 9-year-olds was significant in the critical scenarios ($p < .05$). There were more pronounced age differences in the praise scenarios, in which both 5-year-olds and adults differed from all the other groups ($p < .001$). The multivariate test for the Type × Age interaction was nonsignificant. In addition, there was a significant three-way interaction among valence, type, and age, Wilks's $\Lambda = .70$, $F(3, 92) = 13.18$, $p < .001$, multivariate $\eta^2 = .30$. The pairwise simple-effect comparisons revealed the following differences between the groups: Although the adults judged HC and CP as significantly less mean than all child groups did ($p < .001$), the 5-year-olds judged CP as significantly more mean than both 7- and 9-year-olds ($p < .01$) and HP as significantly less mean than all the other groups ($p < .01$; see Figure 2c and d).

Consistent with the findings on the niceness scale, the participants judged HCs as the meanest statements of all and HP as the least mean of all statements, counterfactual statements being judged in between. Again, the youngest group tended to judge the statement as significantly more mean than the other groups did and the adults judged them as significantly less mean than the children did. Thus, although the youngest group lagged behind the rest of the groups in their judgment of the meanness of irony, the adults were on the other extreme of the scale. Overall, there seems to be a pattern of the decreasing judgment of the meanness of irony with increasing age.

Funniness. Similar to the judgment of niceness, the age effect on the judgment of the funniness of the statements was not significant, $F(3, 92) = 1.82$, $p = .15$, partial $\eta^2 = .06$. However, univariate tests revealed a significant main effect of valence, Wilks's $\Lambda = .71$, $F(1, 92) = 37.99$, $p < .001$, partial $\eta^2 = .29$, and a significant main effect of type, Wilks's $\Lambda = .86$, $F(1, 92) = 15.64$, $p < .001$, partial $\eta^2 = .15$. Praise was judged as significantly funnier than criticism and hyperbole as significantly funnier than counterfactual statements. The interaction between valence and type was also significant, Wilks's $\Lambda = .95$, $F(1, 92) = 4.46$, $p < .05$, multivariate $\eta^2 = .05$: HP was judged funnier than both kinds of counterfactual statements and HC as the least funny type of statement.

Multivariate tests revealed a nonsignificant interaction between type and age but showed a significant interaction between valence and age, Wilks's $\Lambda = .91$, $F(3, 92) = 2.87$, $p < .05$, multivariate $\eta^2 = .07$. The pairwise simple-effect comparisons yielded significant differences in the praise scenarios only, namely, between adults and 5-year-olds ($p < .01$) and 9-year-olds ($p < .05$). The three-way interaction among valence, type, and age was nonsignificant, as were the remaining interactions. There was general consistency among all the child groups and adults in how funny they judged the statements to be. Overall, and contrary to our expectation, not even adults judged the ironic statements as very funny (see Figure 2e and f).

In sum, the age differences in the participants' judgment of the niceness, meanness, and funniness of irony were far less pronounced than was the case for their judgment of the social-cognitive aspects of an ironist's mind. In general, the judgment of irony was not too positive. Yet the adults judged CP as significantly nicer than did the other groups. There was also a trend to regard irony as less mean with increasing age. In support of our hypothesis, ironic criticism was judged as significantly less nice and meaner than ironic praise. By the same token, counterfactual statements were judged less nice and meaner than hyperbole. However, and contrary to the prediction, hyperbole was judged as funnier than counterfactual statements. Overall though, irony was not perceived to be funny by the participants of the study. Thus, both valence and type of statements were significant factors predicting the performance of our groups on their judgment of the pragmatic function of irony.

Discussion

Findings of this study indicate an age-related improvement in children's understanding of the social-cognitive aspects of irony interpretation. Overall, 5-year-olds lagged behind older children in reasoning about the social-cognitive aspects of irony (i.e., speaker's meaning, belief, intention, and motivation), and children of all ages performed more poorly than adults. In contrast, the judgment of the social-communicative functions of irony fell short of displaying a similar developmental trend; the 7- and 9-year-olds judged the niceness and funniness of irony comparably with adults. (Reasons for this discrepancy in trend are discussed later.)

Further, the results offer partial confirmation and partial disconfirmation of the hypothesized asymmetry in understanding different kinds of discourse irony. Understanding counterfactual statements was easier than hyperbole on the belief, intention, and motivation questions. However, the effect of valence (criticism vs. praise) was not significant overall. Although the asymmetry between criticism and praise was revealed within counterfactual statements, confirming Hancock et al.'s (2000) findings, the bias toward criticism or praise was not confirmed in the case of hyperbole (discussed in the following section). In general, ironic messages were not evaluated positively, but adults judged CP more favorably than children did. There was suggestive evidence that ironic messages seemed less mean, although not necessarily nicer, with increasing age. Ironic criticism was considered less nice and meaner than ironic praise, and counterfactual statements were considered less nice and meaner than hyperbolic statements.

Overall, the ironic messages used in the study were not evaluated as humorous. The degree of the judged ''funniness'' of ironic statements varies across different studies. We speculate that the statements used in our study may have been judged as less funny than in some other studies because the judgment of funniness came at the very end of the questionnaire by which time the spontaneity of a potential appreciation of humor might have dissipated in the process of reflecting on the ironist's mind. In addition, the type of irony used will determine the level of funniness; that is, a sarcastic remark would not be expected to be overly funny, nor would an exaggerated criticism. Both of these would likely be judged as more mean and consequently, not too entertaining.

The Role of Valence (Criticism vs. Praise)

The findings revealed no overall effect of valence on the participants' understanding of the social-cognitive aspects of irony. Although the present findings confirm the bias toward an easier understanding of criticism in counterfactual statements (Hancock et al., 2000), the bias was not found in hyperbolic statements. Thus, the present study revealed no general asymmetry between ironic forms of criticism versus praise that would also be expected on the basis of allusional-pretense theory (Kumon-Nakamura et al., 1995). According to the allusional-pretense model, the positive antecedent in the criticism condition is implicitly assumed, as ironic criticism echoes or alludes to presupposed positive norms or expectations. In contrast, ironic praise does not allude to such positive assumptions. For instance, commenting on someone's accomplishment by using a critical statement does not presuppose general negative expectations. As a result, the ironist's explicit echoing of an earlier comment improves the rate of irony detection in the praise condition only (Hancock et al., 2000). The present results are inconsistent with the general assumption that social and cultural norms favor indirect criticism or indirect praise.

In contrast to Hancock et al.'s (2000) design, the present study did not use echoic markers on ironic statements because the types of statements used do not require echoic markers (Gibbs, 2000) and because the participants in the study were older than those in Hancock et al.'s study. Nevertheless, the implicit allusion to positive norms did not help the children in our study to identify the speaker's mental state more readily in critical hyperbolic forms, as allusional-pretense theory would suggest.

Note, however that all forms of irony used (i.e., jocularity as an instance of CP, sarcasm as an instance of CC, and hyperbole) are common in everyday adult conversation (Gibbs, 2000). If children are exposed to these utterances in social interactions, they may accept them as socially appropriate and as alluding to cultural norms in their community. Moreover, the mean age of children in the present study was higher than that in Hancock et al.'s study, making it likely that the school-aged children in our study were better primed to hear both ironic criticism and ironic praise. Thus, the high frequency of these utterance types in everyday discourse and the older age of participants might have reduced potential asymmetries in children's understanding of non-counterfactual criticism versus praise.

Although the valence did not affect children's overall identification of different aspects of the speaker's mind, it affected their evaluation of the pragmatic function of irony. Specifically, older children as well as adults judged ironic praise more positively than ironic criticism; ironic criticism was judged to be considerably meaner than ironic praise, and ironic praise funnier than ironic criticism. Thus, although valence had no overall effect on children's judgment of the speaker's mind, it had a predictably strong effect on the evaluation of the pragmatic function of ironic praise versus criticism.

The Role of Type (Counterfactual vs. Hyperbolic)

The type of irony affected identification of the speaker's beliefs, intentions, and motivations. Specifically, performance on counterfactual statements was superior to that on hyperbolic statements ($p < .001$ on the belief, intention, and motivation variables). This result is consistent with children's recognition of context–utterance discrepancy when the discrepancy is large (e.g., sarcasm) rather than small (e.g., hyperbole; Demorest et al., 1983). The findings are consistent with the claim of Demorest et al. (1983) that recognition of the nonliteral nature of a discrepant utterance (i.e., counterfactual statement) influences the effort to resolve the discrepancy. Because discrepancy and dissonance are greater in counterfactual statements than in hyperbole, the motivation to resolve the discrepancy would be greater (Demorest et al., 1983).

Nevertheless, the present findings contrast with those of Creusere (2000), who found that the form of statements—counterfactual assertions, true assertions, questions, offerings, or thanking—did not affect 8-year-olds' ability to identify the speaker's belief and intention. In other words, in her study children's performance on the counterfactual statements did not differ from that on noncounterfactual statements. Perhaps Creusere's 8-year-olds' failure to identify an ironist's first-order belief and second-order intention more readily in counterfactual statements than in other utterance types stems from ceiling effects. The understanding of first-order beliefs occurs at around 4 years of age, and the understanding of second-order intentions shortly thereafter (Leekam, 1991). Thus, Creusere's findings do not undermine those of the present study, which are consistent with those of Demorest et al. (1983).

Furthermore, the type of utterance had comparably strong effects on children's and adults' judgment of niceness, meanness, and funniness. Specifically, hyperbolic statements were judged as significantly nicer than counterfactual statements, which, in turn, were judged as meaner and less funny than hyperbolic statements.

The findings of this study revealed that understanding counterfactual statements is easier than understanding hyperbole. There are two plausible explanations for this asymmetry: First, children may have greater familiarity with counterfactual statements because jocularity and sarcasm occur more frequently in casual conversations than hyperbole does (Gibbs, 2000). Second, because counterfactual statements signal the discrepancy between context and utterance more strongly than hyperbole, attempts to resolve the dissonance may be more prominent for counterfactual statements than for hyperbole (Demorest et al., 1983). Thus, irony type makes a greater contribution to school-aged children's overall interpretation of irony than the valence of an ironic statement (criticism vs. praise).

Understanding of Social-Cognitive and Social-Communicative Aspects

Consistent with previously published data, our findings demonstrate that although 5-year-olds fail to interpret both the ironist's mind and the communicative effect of irony, 7- and 9-year-olds approach the adult level of interpretation of the social-communicative aspects of some forms of irony. Yet, the novelty of this study is the finding that although school-aged children are able to judge the pragmatic function or the communicative effect of counterfactual and hyperbolic statements in a similar way to adults, they are still unable to properly assess the ironic speaker's mind. Social-cognitive skills at representing beliefs, communicative intentions, and motivations of an indirect speaker continue to develop in and beyond the middle-school years.

Thus, the findings demonstrate continued development during the school-age years in understanding different aspects of mind beyond the well-documented understanding of second-order beliefs (Astington, Pelletier, & Homer, 2002; Perner & Wimmer, 1985; Sullivan et al., 1995). Representing communicative intentions and identifying a speaker's motivations and attitudes are critical aspects of sophisticated communication (Gibbs, 2000; Happé, 1993; Leggitt & Gibbs, 2000; Sperber & Wilson, 1995), particularly in communicative exchanges involving the speaker's allusion to prior attributed thoughts or beliefs (Kumon-Nakamura et al., 1995). The present findings confirm that reflective interpretation of everyday speech is difficult, even for 9-year-olds, who did not reach adult skill levels. In particular, the ability to represent communicative intentions and to identify the motivations and attitudes underlying the speaker's choice of particular utterances develops after the early school years.

In this study, adults in the comparison group were superior to children in constructing an appropriate interpretation of the speaker's intention, motivation, and attitude (attitude coded as the highest score on the motivation coding scheme). Children's interpretation of these aspects of the speaker's mind is imperfect (Pexman & Glenwright, 2007); they have a long way to go before they can construct interpretations of the speaker's complex mental states in the way that adults do (Filippova & Astington, 2008). Importantly, however, we did not find the same striking age differences in understanding of the social-communicative aspects of irony, specifically, in the interpretation of the function of the utterances in discourse (i.e., how nice, mean, and funny the statements are). It is intriguing to speculate about why children find it so easy to interpret the function of some ironic utterances, yet so difficult to infer complex mental states. What distinguishes children's ability to interpret the function of statements from their representation of the speaker's intentions, motivations, and attitudes? How is the process of interpreting the pragmatic function of statements different from that of recognizing a speaker's communicative intent, motivations, and attitudes?

One way to answer these questions is to view the judgment of the speaker's intentions, motivations, and attitudes as requiring interpretation in contrast to the more evaluative character of judging the pragmatic function of ironic statements. That is to say, the assessment of pragmatic function involves one's own evaluation of the communicative effect of irony in the discourse, which is not contingent on one's constructing a mental representation of another's state of mind. In contrast, understanding the social-cognitive aspect of irony depends on making complex interpretative judgments of a speaker's mental states.

This interpretative judgment is a reflective process. For Wilson (2000), the basic metacommunicative capacity is intuitive rather than reflective. Thus, children's ability to assess the pragmatic (i.e., communicative) function of ironic utterances may be more intuitive in the sense that very young children judge things or situations as nice, mean, or funny. Because of exposure to social standards and norms very early in life, children's evaluation of the appropriateness of certain behaviors, including speech, can be construed as intuitive rather than reflective. In contrast, the interpretation of complex intentions, motivations, and attitudes is not a matter of simple intuition. Instead, it involves an active construction of a representation of the speaker's mind (i.e., interpretative judgment). Thus, children's difficulty in interpreting complex mental states may stem from the complex metarepresentational and metacommunicative nature of higher order intentions, motivations, and attitudes. The judgment of the niceness, meanness, and funniness of statements does not require such interpretative skills. Rather, assessing such statements in discourse is less reflective in the sense that this evaluation does not necessitate the representation of cognitive states whereas, in contrast, metarepresentational reasoning is involved in inferences about the speaker's state of mind.

It is also plausible to speculate that the reflective/interpretative and intuitive/evaluative processes take place in different parts of human brain and reveal ontogenetic constraints associated with the development of respective brain regions. In her review of neuroimaging research conducted on theory of mind and empathy, Singer (2006) argues that although our ability to share feelings, emotions, or sensations of others relies on sensorimotor cortices and limbic and paralimbic structures developed early in ontogeny, theory of mind relies on structures of the temporal lobe and prefrontal cortex, maturing considerably later. Singer proposes that the human ability to understand other people's minds and to share affects display different ontogenetic trajectories reflecting the different routes of their underlying neural structures. If the assessment of niceness, meanness, and funniness of irony is a proxy for an empathizing ability (i.e., our participants evaluated the pragmatic impact of the utterances on the addressee from their own perspective), this may help to explain our child participants' ease at evaluating the social-communicative aspects of irony in contrast to their persisting difficulties in judging social-cognitive aspects of irony when compared with the skills documented in adults. As the brain structures associated with understanding others' minds are ones among the last to fully mature (Singer, 2006), it is perhaps not surprising to see the development in understanding complex aspects of a speaker's mind lagging behind the skills at assessing the pragmatic function (i.e., emotional impact) documented to be well under way.

References

Ackerman, B. P. (1983). Form and function in children's understanding of ironic utterances. *Journal of Experimental Child Psychology*, *35*, 487–508.

Andrews, J., Rosenblatt, E., Malkus, U., Gardner, H., & Winner, E. (1986). Children's abilities to distinguish metaphoric and ironic utterances from mistakes and lies. *Communication & Cognition*, *19*, 281–297.

Astington, J. W., Pelletier, J., & Homer, B. (2002). Theory of mind and epistemological development: The relation between children's second-order false-belief understanding and their ability to reason about evidence. *New Ideas in Psychology*, *20*, 131–144.

Clark, H. H., & Gerrig, R. J. (1984). On the pretense theory of irony. *Journal of Experimental Psychology: General*, *113*, 121–126.

Creusere, M. A. (2000). A developmental test of theoretical perspectives on the understanding of verbal irony: Children's recognition of allusion and pragmatic insincerity. *Metaphor and Symbol*, *15*, 29–45.

Demorest, A., Meyer, C., Phelps, E., Gardner, H., & Winner, E. (1984). Words speak louder than actions: Understanding deliberately false remarks. *Child Development*, *55*, 1527–1534.

Demorest, A., Silberstein, L., Gardner, H., & Winner, E. (1983). Telling it as it isn't: Children's understanding of figurative language. *British Journal of Developmental Psychology*, *1*, 121–134.

Dews, S., Winner, E., Kaplan, J., Rosenblatt, E., Hunt, M., Lim, K., et al. (1996). Children's understanding of the meaning and functions of verbal irony. *Child Development*, *67*, 3071–3085.

Filippova, E., & Astington, J. W. (2008). Further development in social reasoning revealed in discourse irony understanding. *Child Development*, *79*, 126–138.

Gibbs, R. W. Jr. (2000). Irony talk among friends. *Metaphor and Symbol*, *15*, 5–27.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Vol. 3. Speech acts* (pp. 41–58). New York: Academic Press.

Hancock, J. T., Dunham, P. J., & Purdy, K. (2000). Children's comprehension of critical and complimentary forms of verbal irony. *Journal of Cognition and Development*, *1*, 227–248.

Happé, F. G. E. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition*, *48*, 101–119.

Happé, F. G. E. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development*, *66*, 843–855.

Harris, M., & Pexman, P. M. (2003). Children's perceptions of the social functions of verbal irony. *Discourse Processes*, *36*, 147–165.

Kreuz, R., & Glucksberg, S. (1989). How to be sarcastic: The reminder theory of verbal irony. *Journal of Experimental Psychology: General*, *118*, 347–386.

Kumon-Nakamura, S., Glucksberg, S., & Brown, M. (1995). How about another piece of pie: The allusional pretense theory of discourse irony. *Journal of Experimental Psychology: General*, *124*, 3–21.

Leekam, S. R. (1991). Jokes and lies: Children's understanding of intentional falsehood. In A. Whiten (Ed.), *Natural theories of mind: Evolution, development and simulation of everyday mindreading* (pp. 159–174). Cambridge, MA: Blackwell.

Leggitt, J. S., & Gibbs, R. W. Jr (2000). Emotional reactions to verbal irony. *Discourse Processes*, *29*, 1–24.

Perner, J., & Wimmer, H. (1985). "John thinks that Mary thinks that…" Attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology*, *39*, 437–471.

Pexman, P. M., & Glenwright, M. (2007). How do typically developing children grasp the meaning of verbal irony? *Journal of Neurolinguistics*, *20*, 178–196.

Pexman, P. M., Glenwright, M., Krol, A., & James, T. (2005). An acquired taste: Children's perceptions of humor and teasing in verbal irony. *Discourse Processes*, *40*, 259–288.

Searle, J. R. (1979). Literal meaning. In J. R. Searle (Ed.), *Expression and meaning: Studies in the theory of speech acts* (pp. 117–136). Cambridge, UK: Cambridge University Press.

Singer, T. (2006). The neuronal basis and ontogeny of empathy and mind reading: Review of literature and implications for future research. *Neuroscience and Biobehavioral Reviews*, *30*, 855–863.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Cambridge, MA: Blackwell.

Sullivan, K., Winner, E., & Hopfield, N. (1995). How children tell a lie from a joke: The role of second-order mental state attributions. *British Journal of Developmental Psychology*, *13*, 191–204.

Wilson, D. (2000). Metarepresentation in linguistic communication. In D. Sperber (Ed.), *Metarepresentations: A multidisciplinary perspective* (pp. 411–448). New York: Oxford University Press.

Winner, E., & Leekam, S. (1991). Distinguishing irony from deception: Understanding the speaker's second-order intention. *British Journal of Developmental Psychology, 9*, 257–270.

Winner, E., Windmueller, G., Rosenblatt, E., Bosco, L., Best, E., & Gardner, H. (1987). Making sense of literal and nonliteral falsehood. *Metaphor and Symbolic Processes, 2*, 13–32.

Appendix

Sample of Irony Task

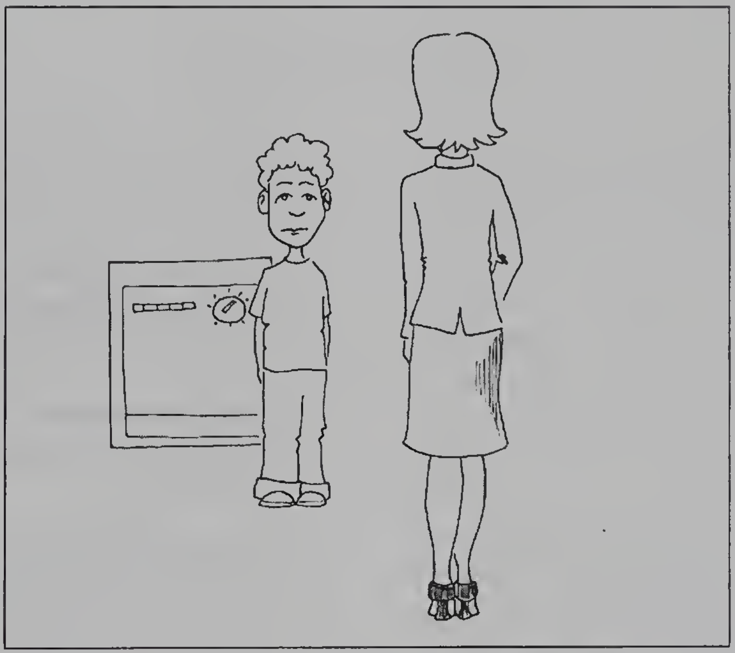
Dinner
Mom asks Billy to help her empty the dishwasher before dinner.
Billy starts to empty the dishwasher while mom is making dinner.

Criticism:
Billy drops a plate and it breaks.
Mom turns around and says to Billy:

Counterfactual: "You sure ARE a GREAT helper!"
Hyperbolic: "You sure KNOW how to WRECK the house!"

Praise:
Billy neatly puts all the dishes from the dishwasher in the cupboards.
He even sets the table for dinner.
Mom turns around and says to Billy:

Counterfactual: "You sure ARE a CRUMMY helper!"
Hyperbolic: "You MUST be a kitchen SUPERMAN!"



Sample of Questionnaire

Counterfactual Criticism Condition

Comprehension Questions:				
1. Did Billy end up helping his mom?	[correct answer: no]			
2. What did mom say to Billy?	[exact retell, a paraphrase, or the correct gist]			
[exact statement reiterated: "*You sure ARE a GREAT helper!*"]				
Meaning:				
3.1 Does mom mean that?	[correct answer: no]			
3.2 What does she mean?	[acknowledged meaning opposite to the literal statement]			
Belief:				
4. Does mom think Billy is a great helper?	[correct answer: no]			
Intention:				
Mom said to Billy: "*You sure ARE a GREAT helper!*"				
5. Does mom want Billy to believe that she thinks that?	[correct answer: no]			
Motivation/Attitude:				
Mom said to Billy: "*You sure ARE a GREAT helper!*"				
6. Why did she say that?	[coded on a 5-point scale]			
Pragmatic Function:				
7. Mom said to Billy: "*You sure ARE a GREAT helper!*"				
Pick one of these four glasses to show				
(a) How nice was what mom said?	0	1	2	3
(b) How mean was what mom said?	0	1	2	3
(c) How funny was what mom said?	0	1	2	3

Sample of Questionnaire

Hyperbolic Criticism Condition

Comprehension Questions:				
1. Did Billy end up helping his mom?	[correct answer: no]			
2. What did mom say to Billy?	[exact retell, a paraphrase, or the correct gist]			
[exact statement reiterated: "*You sure KNOW how to WRECK the house!*"]				
Meaning:				
3.1 Does mom mean that?	[correct answer: no]			
3.2 What does she mean?	[acknowledgement of meaning reduced from the exaggeration]			
Belief:				
4. Does mom think Billy knows how to wreck the house?	[correct answer: no]			
Intention:				
Mom said to Billy: "*You sure KNOW how to WRECK the house!*"				
5. Does mom want Billy to believe that she thinks that?	[correct answer: no]			
Motivation/Attitude:				
Mom said to Billy: "*You sure KNOW how to WRECK the house!*"				
6. Why did she say that?	[coded on a 5-point scale]			
Pragmatic Function:				
7. The mom said to Billy: "*You sure KNOW how to WRECK the house!*"				
Pick one of these four glasses on the picture to show				
(a) How nice was what mom said?	0	1	2	3
(b) How mean was what mom said?	0	1	2	3
(c) How funny was what mom said?	0	1	2	3

Child Development, May/June 2010, Volume 81, Number 3, Pages 929–944

Inconsistency With Prior Knowledge Triggers Children's Causal Explanatory Reasoning

Cristine H. Legare
University of Texas at Austin

Susan A. Gelman and
Henry M. Wellman
University of Michigan

What events trigger causal explanatory reasoning in young children? Children's explanations could be triggered by either consistent events (suggesting that explanations serve a confirmatory function) or inconsistent events (suggesting that they promote discovery of new information). In 2 studies with preschool children ($N = 80$), events that were consistent with children's prior knowledge were simultaneously contrasted with events that were inconsistent with prior knowledge, and children were invited to explain either outcome (or both). Results demonstrate that inconsistent outcomes are an especially powerful trigger for children's explanations and that the explanations children provide for inconsistent outcomes refer to internal causal properties, overriding perceptual appearances. In sum, the data provide empirical evidence that inconsistent events motivate children to construct explanations, thereby suggesting that children's explanations function in the service of discovery.

> *Observation is always selection. It needs a chosen object, a definite task, an interest, a point of view, a problem.*
>
> —Popper, 1963

A fundamental task for all humans is explaining why things happen. By 3 years of age, children can use causal knowledge to make predictions (Shultz, 1982), engage in effective interventions (Kushnir & Gopnik, 2005; Schulz & Gopnik, 2004), and provide explanations for things that happen (Wellman, Hickling, & Schult, 1997). Not only do young children frequently seek explanations by asking questions, they also construct their own explanations (Callanan & Oakes, 1992; Hickling & Wellman, 2001).

At the same time, young children are surprisingly poor at assessing their own explanatory understanding. This ability develops dramatically across development, but adults and especially children overestimate the detail and depth of their explanatory knowledge (Mills & Keil, 2004; Wilson & Keil, 1998). Thus, on the one hand, children are active explanation seekers and readily request and provide causal explanations. Yet on the other hand, they seem to be poor at assessing their own causal knowledge and often think they understand things when they do not. What then motivates children to ask questions and generate explanations? More specifically, what kinds of events provoke causal explanatory reasoning in children? Children could make use of causal-explanatory understanding when reasoning about either *consistent events* (when events unfold as anticipated based on prior knowledge) or *inconsistent events* (when something unusual or contrary to prior knowledge happens).

Accordingly, children's explanations may serve at least two distinct functions. First, explanation might be systematically confirmatory. Research on the development of scientific reasoning provides evidence for a confirmation bias in both children and adults (Nickerson, 1998) whereby reasoners tend to seek information that supports a favored hypothesis and to interpret events in ways that are partial to that hypothesis. Both children and adults also tend to avoid, ignore, or distort information that is inconsistent with their beliefs (Kuhn, 1989).

This research was supported by a foundation grant from the McDonnell foundation to Wellman and Gelman, NICHD Grant HD22149 to Wellman, and NICHD Grant HD36043 to Gelman. We are grateful to the teachers, staff, and children of the University of Michigan Children's Centers, Pooh Corner Preschool, and Bemis Farms Preschool for participating in this research. We would also like to thank Tamar Kushnir for her insightful input on study design, Paul Theron for his assistance with stimuli design, and Andrea Kiss for her assistance with data collection.

Correspondence concerning this article should be addressed to Cristine H. Legare, Department of Psychology, University of Texas at Austin, 1 University Station A8000, Austin, TX 78712-0187. Electronic mail may be sent to legare@psy.utexas.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8103-0019

Confirmatory information can be favored and privileged even when people have no vested interest in a hypothesis (Bruner, Goodnow, & Austin, 1956).

There is much evidence for a confirmation bias in older children and adults. But young children may be especially prone to such a bias, because they are early in the process of developing explanatory knowledge and so can benefit from consolidating, confirming, and deepening their current explanations. Thus, not only do young children anticipate that future outcomes will proceed in accord with prior expectations (Gopnik, Sobel, Schulz, & Glymour, 2001; Kushnir & Gopnik, 2007; Schulz & Gopnik, 2004), they may also find such outcomes easy to explain and worthy of explanation.

Alternatively, however, young children's explanation might be especially triggered by *inconsistent* information, that is information inconsistent with expectations. If so, providing explanations might be as much a childhood process of discovery as a process of confirmation. Children may systematically attend to and attempt to explain information that is inconsistent with expectations. Notably, the motivation to construct causal explanation may be driven by the puzzle of inconsistent outcomes, and not merely by unpredicted, low probability, or novel outcomes. Forming explanations for inconsistent outcomes may provide children with the opportunity to generate new hypotheses, and to accommodate inconsistent information in the context of prior beliefs.

Several arguments and findings make this alternative possibility plausible as well. The idea that inconsistent, problematic, or surprising outcomes play an important role in causal reasoning for adults and school-age children appears across multiple literatures—including philosophy of science (Hempel, 1965), social psychology (Hilton, 1995; Weiner, 1985), and educational research (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Wong & Weiner, 1981). Moreover, it is well documented that infants often attend to inconsistent or unexpected outcomes (Baillargeon, 2002).

Against these theoretical possibilities, there is a dearth of empirical research on what actually motivates young children to construct explanations. Thus, it is unclear which model better predicts children's causal explanations. We take an empirical approach by assessing what motivates young children to provide the explanations that they do.

The present studies focus on children's explanations for confirmatory versus disconfirmatory events. We provided children with the following basic scenario. Children viewed two equivalent events simultaneously: one in accord with prior knowledge and the other not. If, for young children, explanation is largely confirmatory, then they should simply explain what they already have an explanation for. If explanation is instead especially employed for discordant or anomalous information, children should explain the event that falls outside their prior expectations. We measured not only *what* children explain but also *how* they do so, by examining whether they appealed to surface features or instead to less obvious causal properties and category membership. We assume that explanations reflect the nature of children's conceptual understanding (Keil, 2006; Legare, Wellman, & Gelman, 2009; Lombrozo, 2006; Medin, Coley, & Storms, 2003; Sloman, 1994). For example, consider category membership. Even children as young as 3 years of age can categorize objects in terms of novel, nonobvious properties (Gelman & Coley, 1990; Gelman & Markman, 1986, 1987; Graham, Kilbreath, & Welder, 2004; Jaswal & Markman, 2007), apply names to objects with the same functional properties (Kemler-Nelson, 1995), and categorize and name objects based on novel causal properties (Gopnik & Sobel, 2000), overriding perceptual appearances. Whether and how preschool children might make use of such information in their causal explanations (as opposed to their predictions or judgments) is, however, an open question.

Study 1

To address these issues, we designed a set of novel "light boxes"—devices that glowed brightly when activated. The activation of the boxes was experimenter controlled, but appeared to be caused by placing an object on top (modeled after Gopnik & Sobel, 2000; Gopnik et al., 2001). These materials were first used to teach children briefly about different categories of objects that turned a box on, turned it off, or did nothing. In Study 1, these training objects were labeled according to their causal properties: "starters" (which lit the light box when placed on top), "stoppers" (which turned off a lit box), and "do-nothings" (which neither lit nor unlit the light box). Items within each object category (e.g., do-nothings) were perceptually identical. After training, children were presented with scenarios in which a new object that looked like one type (e.g., it looked like a starter) actually behaved like another type (e.g., it behaved like a do-nothing). This was simultaneously paired within test trials where a contrasting object looked and behaved like those previously seen (e.g., it looked like a do-nothing and behaved like a do-nothing). Upon viewing such paired outcomes, children were asked a nonspecific explanatory question ambiguously referring to either (visible) outcome: "Why did that happen?"

Comparing outcomes consistent or inconsistent with prior knowledge was our primary interest, but state changes and negative outcomes may also function as explanatory triggers. Therefore, two focal conditions were designed that controlled for these factors by holding them constant. In the *generative condition*, both light boxes began off, both turned on in one of the test trials, and both remained off in the other test trial. So, in the test trial in which both boxes turned on, both outcomes involved state change and neither outcome involved a negative outcome (defined here as failing to turn the boxes on). In the parallel test trial in which both boxes stayed off, neither outcome involved a state change and both outcomes involved a negative outcome (failing to turn the boxes on). In the contrasting *inhibitory condition*, both light boxes began on and this pattern of outcomes was reversed.

Prior to these test trials but after initial training, children participated in a confirmation trial. The confirmation trial not only served to remind children of how various objects functioned, but also had the advantage of isolating change-of-state outcomes from outcomes inconsistent with prior knowledge. For example, in the confirmation trials, because both outcomes work as expected based on prior experience, starters and stoppers changed the state of the box, and do-nothings did not change the state of the box.

Our interest was in which events yielded not just greater attention, but most importantly, increased explanation. If state change plays a role, on the *confirmation trials* children should differentially explain outcomes when a box turned off (after being on) or turned on (after being off). If negative outcomes provoke explanation, on the *confirmation trials* children should differentially explain outcomes when a box failed to turn on, or when a box turned off. Most importantly, if the role of explanations for children is confirmatory, children should be interested in and provide explanations for the consistent event in each *test trial* pair, whereas if the role of explanations is discovery, children should be specifically interested in and provide hypotheses and explanations for the inconsistent event in each *test trial* pair.

Method

Participants

Sixteen 3-year-olds (*M* age = 3.6; range = 3.2–3.11), sixteen 4-year-olds (*M* = 4.6; range = 4.0–4.11), and sixteen 5-year-olds (*M* = 5.3; range = 5.0–5.6) were recruited from a Midwestern U.S. university town. Participants were primarily White and from middle-class families. Approximately equal numbers of boys and girls participated in the study.

Materials

Four light boxes (5 in. × 5 in. × 5 in.) were made of wood with thin translucent laminate tops. The boxes were identical except one pair was red and the other pair was yellow. Each pair of boxes (red or yellow) was attached to a switch box operated surreptitiously by a confederate. If the switch was turned on, the box would light up and stay on until it was switched off. Alternately, if the switch was turned off, the box would turn off and stay off until it was switched on. For objects with a causal effect, each box was turned off or on as soon as an object made contact with it, and would stay off or on until the object was removed. This yielded the strong impression that some objects turned the boxes on, some turned them off, and some had no effect on the boxes. The switchbox was hidden from the children's view, and none of the participants mentioned the confederate or indicated any suspicion that the confederate influenced the functioning of the boxes.

There were four perceptually distinct sets of objects, shown in Figure 1, each consisting of five identical wooden blocks (20 blocks total). Objects were always presented in paired sets. The perceptual features of different kinds of objects were not indicative of the observed causal properties of the objects (lighting or not lighting the boxes). Additionally, the placement of the objects on the boxes was counterbalanced across trials.

Design

Trials were designed both to isolate state change and negative outcomes as potential explanatory triggers (confirmation trial) and to control for these factors by holding them constant (test trials). To help assess these factors, there were two different conditions. In the *generative condition*, the boxes (either yellow or red, counterbalanced across children) started with the light off, and in the training trial, the two distinctive sorts of objects were

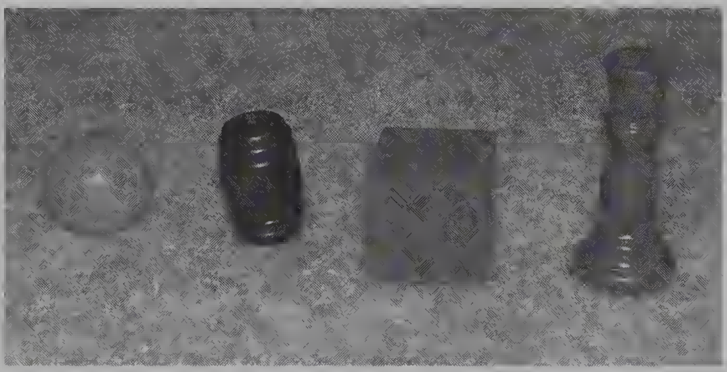


Figure 1. Studies 1 and 2: Pictures of object stimuli.

labeled as either starters or do-nothings. In the *inhibitory condition*, this pattern was reversed; the boxes started with the light on, and, in the training trial, the objects were labeled as either stoppers or do-nothings. One test trial displayed both light boxes turned on and the other test trial displayed both light boxes turned off. Therefore, in each test trial, inconsistency with prior knowledge was isolated; one outcome was inconsistent with prior knowledge and one outcome was consistent with prior knowledge, but both changed (or don't change) state, and both boxes were on (positive) or off (negative).

Generative condition. Table 1 provides an outline of the factors and design for the entire study. In each condition, after training, there was one confirmation trial followed by two test trials. For the generative condition, in the confirmation trial, both outcomes were consistent; there was no inconsistent outcome. As listed in Table 1, the object that looked like a starter turned the box on and therefore produced a state change and a positive outcome (light box was on). The object that looked like a do-nothing did not change the state of the box and produced a negative outcome (light box was off; see Figure 2). Therefore, if the child was motivated to explain a state change, he or she would have explained the outcome involving the object that looked like a starter first, and if the child was motivated to explain the negative outcome, he or she would have explained the object that looked like a do-nothing first.

In the generative test trial in which both boxes turned on, both outcomes involved state change and neither involved a negative outcome (defined here as failing to turn the boxes on). However, the object that looked like a starter produced an outcome consistent with prior knowledge and the object that looked like a do-nothing produced an outcome inconsistent with prior knowledge (see Table 1). Therefore, if the child was motivated to explain inconsistency with prior knowledge, he or she should explain the object that looked like a do-nothing first. In the parallel generative test trial in which both boxes stayed off, neither outcome involved a state-change and both outcomes involved a negative outcome (both failed to turn the boxes on). In this trial, instead of the object that looked like a do-nothing producing an inconsistent outcome, it was the object that looked like a starter that produced an inconsistent outcome.

Inhibitory condition. In the inhibitory confirmation trial, again both outcomes were consistent;

Table 1
Studies 1 and 2: Study Design

Condition	Generative (off → X)						Inhibitory (on → X)					
Trial	Confirmation trial		Test trial		Test trial		Confirmation trial		Test trial		Test trial	
Initial state of light box	Off	Off	Off	Off	Off	Off	On	On	On	On	On	On
Type of object	Starter	Do-nothing	Starter	Do-nothing	Starter	Do-nothing	Stopper	Do-nothing	Stopper	Do-nothing	Stopper	Do-nothing
Final state of light box	On	Off	Off	Off	On	On	Off	On	Off	Off	On	On
Inconsistent outcome = 1	0	0	1	0	0	1	0	0	0	1	1	0
State change = 1	1	0	0	0	1	1	1	0	1	1	0	0
Negative outcome = 1 (off)	0	1	1	1	0	0	1	0	1	1	0	0
Isolates	State change/ negative outcome		Inconsistent outcome		Inconsistent outcome		State change/ negative outcome		Inconsistent outcome		Inconsistent outcome	

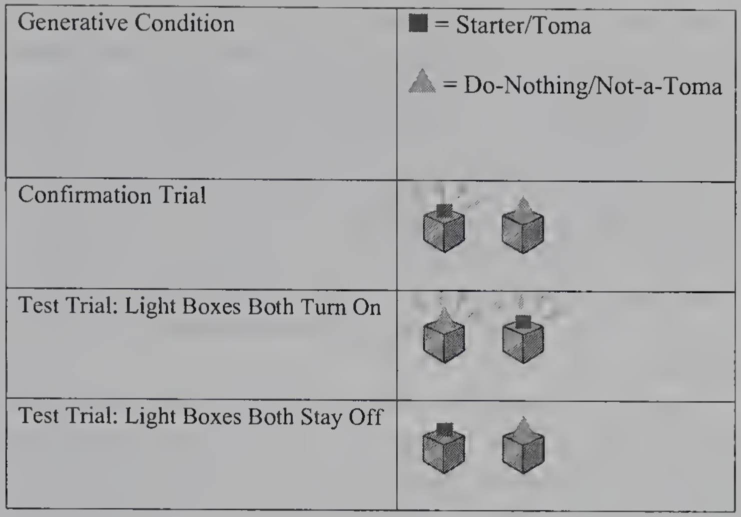

Generative Condition
■ = Starter/Toma
▲ = Do-Nothing/Not-a-Toma
Confirmation Trial
Test Trial: Light Boxes Both Turn On
Test Trial: Light Boxes Both Stay Off

Figure 2. Studies 1 and 2: Generative condition.

there was no inconsistent outcome. As seen in Table 1, the object that looked like a stopper turned the box off and therefore produced a state change *and* a negative outcome (the light box was off). This correspondence of a state change and negative outcome was unlike the generative condition, in which state change and negative outcome competed in the confirmation trial. The object that looked like a do-nothing did not change the state of the box and produced a positive outcome (the light box was on; see Figure 3). So if the child was motivated to explain a state change or a negative outcome, he or she would have explained the outcome involving the object that looked like a stopper first. If neither state change nor negative outcomes were explanatory triggers, the child would have explained the object that looked like a do-nothing first.

In the inhibitory test trial in which both boxes turned on, neither outcome involved state change nor a negative outcome. However, the object that looked like a do-nothing produced an outcome consistent with prior knowledge, and the object that looked like a stopper produced an outcome inconsistent with prior knowledge (see Table 1). Therefore, if the child was motivated to explain inconsistency with prior knowledge, he or she should explain the object that looked like a stopper first (the opposite pattern of the generative condition in the test trial, in which both boxes lit up). In the parallel inhibitory test trial in which both boxes turned off, both outcomes involved a state-change and both outcomes involved a negative outcome (both turned the boxes off). In this trial, instead of the object that looked like a stopper producing an inconsistent outcome, it was the object that looked like a do-nothing that produced an inconsistent outcome.

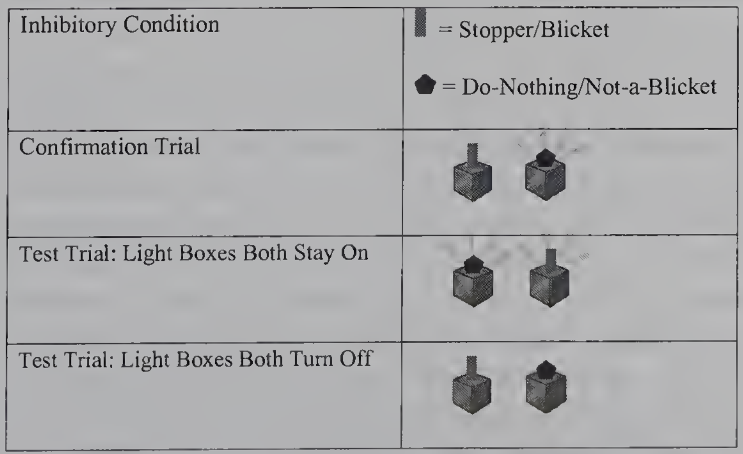

Inhibitory Condition
▮ = Stopper/Blicket
⬟ = Do-Nothing/Not-a-Blicket
Confirmation Trial
Test Trial: Light Boxes Both Stay On
Test Trial: Light Boxes Both Turn Off

Figure 3. Studies 1 and 2: Inhibitory condition.

Procedure

Children were tested individually in a 15-min session seated across from the experimenter and confederate at a table on which a pair of the wooden boxes was placed. Each child participated in both conditions: generative and inhibitory. After the objects' functions and labels were presented individually in the training trials, both objects were presented simultaneously on three trials: one confirmation trial followed by two test trials. Although the objects in the confirmation and test trials were perceptually identical to the objects in the training trial, they were never labeled for the child. For example, in the generative condition, after demonstrating how a starter and a do-nothing work for one set of objects in the training trial, the experimenter placed another object perceptually identical to a starter and another object perceptually identical to a do-nothing simultaneously on two unlit boxes. In this confirmation trial, both objects worked in a manner consistent with prior knowledge (upon contact with one box, the object that looked like a starter turned the light on, and the object that looked like a do-nothing had no effect). Children were asked to explain this initial paired event, to provide a baseline comparison for their subsequent test trial explanations. Specifically, with both objects on their boxes, one box on and one off, children were asked, "Why did that happen?"

For each test trial, two objects were again simultaneously placed on the boxes. One object worked in a manner consistent with prior knowledge and the other object worked in a manner that was inconsistent. Unlike the confirmation trial, in the test trials the outcomes in both cases were causally identical (both boxes lit up in one test trial and neither box lit up in the other test trial). After viewing the paired outcomes of each trial, children were again asked, "Why did that happen?" Children were prompted several times in a nondirected manner after that initial causal question to provide additional explanatory information (e.g., "Can you tell me more?" or "Do you have any other ideas?"). For each trial, the child was encouraged to provide as many explanations as possible, and the trial was terminated after the child indicated that he or she had no other information to share.

Each child had one generative block of trials with one pair of light boxes (one confirmation trial and two test trials) and one inhibitory block of trials with the other pair of light boxes (one confirmation trial and two test trials). Which condition came first (generative or inhibitory) was counterbalanced.

Transcription and Coding

Interviews were videotaped and transcribed verbatim yielding the following measures for each trial: what the child looked at first, what he or she explained first, his or her overall explanatory response, and the kind of explanations he or she provided. During the confirmation trial, the outcome the child looked at first was coded (starter or stopper = 1, do-nothing = 0), followed by the outcome the child mentioned first (starter or stopper = 1, do-nothing = 0, both = 0.5, if a child referred to both outcomes in their first explanation).

For each test trial per condition, as in the confirmation trial, the outcome the child looked at first was coded (inconsistent = 1, consistent = 0). The outcome the child explained first was coded (inconsistent = 1, consistent = 0, both = 0.5), followed by a similar coding (inconsistent, consistent) for any additional explanations the child provided. Children's explanations for inconsistent and consistent outcomes were coded separately. For example, if the child provided an explanation for the inconsistent outcome, "'Cause this starter is broken, the parts inside don't work," and then provided an explanation for the consistent outcome, "'Cause this do-nothing is working, it doesn't make anything happen to the box," the child was coded as explaining the inconsistent outcome first, and then their overall explanations were coded individually as providing at least one explanation for the inconsistent outcome and providing at least one explanation for the consistent outcome.

Explanatory Categories

Children's explanations were also coded for content, separately for consistent and inconsistent outcomes. Causal explanations for inconsistent outcomes were coded into three primary categories: category switch, causal function, and causal action. In category switch explanations, children referred to a switch in category membership based on function (e.g., "It is a do-nothing now" or "It's really a starter; it only looks like a do-nothing"; see Table 2 for more examples). Notably, such explanations go beyond the past history of perceptually identical objects.

Explanations that discussed a problem with the functioning of the object were coded as causal function explanations. These explanations included reference to the object being broken ("The stopper is not working anymore. It is broken"), the box being broken ("The box does not light anymore. It is broken"), or differences among the objects ("This one is heavier than the others"). Explanations that referred to insides were also included ("There aren't stoppers inside" or "All out of batteries"). Explanations that referred to problems using the objects or problems with the placement of the objects were coded as causal action explanations ("You set the stopper/do nothing on the wrong sides" or "It's on the wrong box"; see Table 2).

Noncausal explanations were coded into four other categories: expectation violation, descriptive statements, psychological references, and don't know. Explanations that described what could be expected to happen on the basis of appearance or past events without providing a cause were coded as expectation violation (e.g., "it wasn't supposed to turn on"). Explanations that referred to the criteria children used as a basis for their conclusion without further explanation were coded as *descriptive statements* (e.g., "It's not on because it's not glowing up"). Explanations that referred to psychological constructs such as preference or desire (e.g., "maybe it wanted to") were also recorded, although these were quite rare. If children were unable to provide an explanation, their responses were coded as *don't know.*

Causal explanations for consistent outcomes were coded into three primary parallel categories: category label, causal function, and causal action. Explanations that referred to category membership using a label were coded as *category label* explanations (e.g., "Because this one is a stopper" or "this one is a do-nothing"). Unlike the category switch explanations for inconsistent outcomes, category label explanations involved providing a label that was consistent with the past history of perceptually identical objects. Causal function, causal action, and noncausal explanations for consistent outcomes

Table 2
Sample Causal Explanations for Inconsistent Outcomes

Category switch (Study 1): "It looks like there was one do-nothing, and it seems like it wasn't going to start but it did start. It's really a starter." "It looks like that one was a do-nothing, but it wasn't a do-nothing it was a stopper. It makes the box stop." "Because that is a starter, too. Because some are starters, too. Some of these are starters too."

Causal function (Study 1): "The stopper isn't working anymore; it's broken inside." "Because this went out of power. Because it doesn't have stoppers in it." "Because it's broken. I think because it doesn't have batteries in it."

Causal action (Study 1): "You set the starter on the wrong side." "The stopper should be on the other box, that's why it isn't working right." "They are on the wrong boxes."

Causal category (Study 2): "It's a toma, it just looks like it isn't one." "It turned on because it is a blick." "Because that one can make it turn off. It is the same kind of thing as the toma."

Causal function (Study 2): "The toma isn't working anymore; it's broken inside." "Maybe because this one has more air and energy." "The toma didn't stop. It didn't stop because it doesn't have any magnets in it."

Causal action (Study 2): "You set the blicket on the wrong side." "The toma should be on the other box; maybe try putting it on there." "Because that one (blicket) is supposed to be on the other side and the one that isn't a blicket is supposed to be on the other box."

were coded using the same criteria as explanations for inconsistent outcomes.

Interrater reliability was established using a randomly selected sample of 25% of the data. Two persons independently coded the reliability sample with 95% agreement. Reliability was also calculated for explanatory coding categories for both consistent and inconsistent outcomes, with kappas ranging from .86 to .94 (near perfect levels; Landis & Koch, 1977).

Results

We examined separately the events that resulted in greater attention and the events that resulted in increased explanation.

Which Outcomes Do Children Look at and Explain First?

Confirmation trials. In the confirmation trials, children were asked to explain a paired event in which both objects functioned in a manner consistent with prior knowledge. Among two possible trials, children were more likely both to look first at the object (starter or stopper) that changed the state of the box (63% of the time), $t(47) = 2.84$, $p < .01$, and to first explain the object that changed the state of the box (71% of the time), $t(47) = 4.57$, $p < .001$, when compared with the object that had no effect on the box (do-nothing). Beyond initial attention, children were significantly more likely to first explain the object that changed the state of the box in both the generative ($M = 69\%$) and the inhibitory ($M = 74\%$) conditions, with both of these greater than chance, $ps < .001$. Thus, in the absence of an inconsistent event or outcome, state change systematically trigged children's explanations. Negative outcomes, however, did not; even in the generative condition, where the confirmation trial pitted a state change event against a negative outcome event, children were significantly more likely to first explain the state change outcome than the negative outcome.

Test trials. In the test trials, events inconsistent with children's prior knowledge first attracted children's attention and provoked their explanations. Thus, across age groups and test trials (of four possible trials), children were much more likely to look first at the inconsistent outcomes than the consistent outcomes ($M = 3.73$, or 93% of the test trials), $t(47) = 18.61$, $p < .001$. More importantly, children were much more likely to explain the inconsistent outcome first ($M = 3.18$ out of 4, or 80%), $t(47) = 11.72$, $p < .001$.

In order to investigate potential age and condition-related effects, we conducted two separate repeated measures analyses of variance (ANOVAs) involving Age Group (3-, 4-, and 5-year-olds) × Condition (inhibitory vs. generative) × Final Outcome (both boxes on vs. both boxes off). Condition and final outcome were within-subjects factors, and age group was a between-subjects factor. One ANOVA involved first-look responses; the other involved first-explanation responses. In all age groups and conditions, children were very likely to look first at (3-year-olds on 90% of their test trials, 4-year-olds 95%, and 5-year-olds 95%) and explain first (3-year-olds 85%, 4-year-olds 78%, and 5-year-olds 78%) the inconsistent outcome. There was no main effect of age (for either the first-look or first-explanation analysis). Similarly, there were no significant main effects of condition (inhibitory vs. generative) or final outcome (both on vs. both off).

However, for the first-explanation analysis, the Condition × Final Outcome interaction was significant, $F(1, 45) = 4.62$, $p < .05$. Post hoc tests indicated that for the inhibitory condition, children were more likely to explain the inconsistent outcome when both boxes remained on (no state change) (84% of the time) than when both boxes turned off (state change; $M = 70\%$), $p < .05$. Importantly however, in both cases children were significantly more likely to look first at the inconsistent outcome than chance alone would predict, $ps < .001$.

As noted earlier, children participated in a confirmation trial prior to the test trials. One possibility is that participating in a confirmation trial may have influenced the pragmatics of the explanation task. For example, because children provided an explanation for an outcome in the confirmation trial, they may have been motivated to explain a different outcome in the trial that came next (i.e., the first test trial). Thus, if children provided an explanation for a state-change outcome in the confirmation trial (object that looks like a starter turned the box from off to on), and if their first test trial included a perceptually identical outcome (object that looks like a starter turned the box from off to on) compared with a different outcome (object that looks like a do-nothing turned the box from off to on), they may have been motivated to explain the latter because they had just explained the former on the prior (confirmation) trial. To investigate whether this sequencing influenced the results, we compared what the children explained first in their first test trial (following the confirmation trial) with what they explained first in their second test trial. There were no significant differences for test trial order, $t(47) = 0.97$, *ns*.

We also analyzed children's spontaneous explanations (provided prior to an explicit explanatory question) in order to address pragmatic concerns with the procedure. Namely, this analysis allows us to rule out the possibility that children were simply generating the explanation they believed the experimenter was requesting. For first explanations, spontaneous explanations revealed the same pattern of responses as elicited explanations. Sixteen of 48 children (33%) provided a spontaneous explanation at least once in the test trials. For 81% of the spontaneous explanations, children explained the inconsistent outcome first, $F(1, 47) = 132.89$, $p < .001$. Children's spontaneous explanation of inconsistent outcomes provides clear evidence that they were not merely explaining what they thought the experimenter was interested in.

Which Outcomes Do Children Provide an Explanation for Overall?

Children often provided multiple explanations for any test trial: 19%, 25%, and 50% of trials at ages 3, 4, and 5 years, respectively. For this analysis, the total number of explanations provided was the dependent variable. We conducted a 3 (age group: 3-, 4-, and 5-year-olds) × 2 (condition: inhibitory vs. generative) × 2 (final outcome; both boxes on vs. both boxes off) × 2 (consistent vs. inconsistent outcome) repeated measures ANOVA on the entirety of children's explanations. Condition, final outcome, and consistent versus inconsistent outcome were within-subjects factors. As can be seen in Figure 4, there were significant main effects for age group, $F(2, 45) = 14.18$, $p < .001$; condition, $F(1, 45) = 12.71$, $p < .001$; and, most importantly, consistent versus inconsistent outcome, $F(1, 45) = 63.12$, $p < .001$. Post hoc tests indicated that 5- and 4-year-olds gave explanations more consistently than 3-year-olds (Ms = 94%, 89%, and 69% of total responses, respectively), which is not surprising, given developmental changes in children's language skills over this age period. Most importantly, however, the significant main effect for consistent versus inconsistent outcome demonstrates that inconsistent outcomes were more likely to be explained than consistent outcomes (Ms = 3.98 and 2.73 of 4, or 99% and 68%, respectively). Not shown in Figure 4 is that overall explanations were more prevalent in the inhibitory ($M = 88\%$) than the generative condition ($M = 80\%$), as shown by the effect of condition. We can offer no firm account for this effect, but perhaps children found the initial state of the light boxes being on (inhibitory condition) more novel or unexpected than the initial state of the light boxes being off (generative condition).

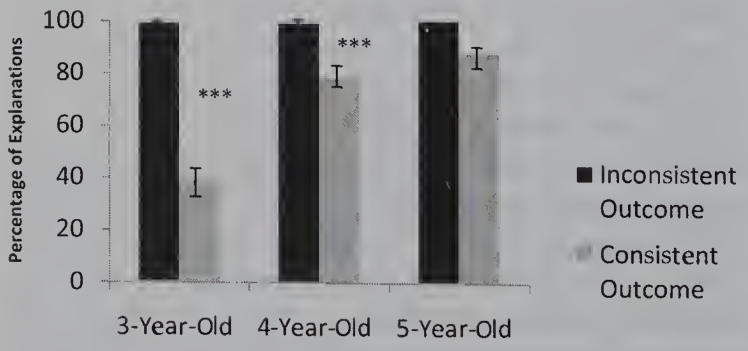


Figure 4. Study 1: Percentage of explanations children provided for inconsistent versus consistent outcomes by age group.
Note. Comparisons between explanations for consistent versus inconsistent outcomes.
$^{***}p < .001$.

The ANOVA also yielded a significant Consistent Versus Inconsistent Outcome × Age Group interaction, $F(2, 45) = 13.90$, $p < .001$, indicating developmental differences in the outcome children were most likely to explain overall. Post hoc tests reveal that although 3- and 4-year-olds were more likely to explain inconsistent than consistent outcomes, $ps < .001$, 5-year-olds were equally likely to explain both. The lack of effect for 5-year-olds reflects a ceiling effect (see Figure 4). There were as well a significant Condition × Consistent Versus Inconsistent Outcome interaction, $F(1, 45) = 12.62$, $p < .001$; a Condition × Final-Outcome interaction, $F(1, 45) = 6.88$, $p < .05$; and a three-way Condition × Final Outcome × Consistent Versus Inconsistent Outcome interaction, $F(1, 45) = 8.25$, $p < .001$. Nonetheless, for all four test trials (in each condition and each final outcome), children were more likely to provide an explanation for an inconsistent than a consistent outcome.

The Kinds of Causal Explanations Children Provided

Recall that children's explanations were coded into three primary causal categories: category switch (or category label in the case of consistent events), causal function, and causal action explanations. The majority of children's explanations at all age groups were causal (rather than the noncausal categories of expectation violation, descriptive statements, psychological references, and don't know). This was true for both inconsistent outcomes (55%, 81%, and 95% of the explanations provided at ages 3, 4, and 5, respectively) and consistent outcomes (76%, 74%, and 88% of the explanations provided at ages 3, 4, and 5). Explanations that made use of psychological explanations were very rare; only one 3-year-old and two 4-year-olds used psychological language in their explanations ("This one thinks it's a starter").

An Age Group (3-, 4-, and 5-year-olds) × Condition (inhibitory, generative) × Consistent Versus Inconsistent Outcome × Causal Explanation Type (causal function, causal action, category label/category switch) repeated measures ANOVA was conducted in order to investigate age and condition-related changes in the kinds of causal explanations children provided. Condition, consistent versus inconsistent outcome, and causal explanation type were within-subject factors. The number of total explanations provided (of each type) was the dependent variable. This analysis yielded the same main effects for age group, $F(2, 45) = 10.83$, $p < .001$, and consistent versus inconsistent outcome, $F(1, 45) = 24.19$, $p < .001$, described earlier; therefore, only results involving the different types of causal explanations given will be discussed here. Here, there was a significant main effect for causal explanation type, $F(2, 90) = 11.64$, $p < .001$. Causal function and category switch explanations (or category label explanations for consistent outcomes) were most common (and not significantly different from one another), and each was more frequent than causal action explanations, $ps < .01$.

Additionally, the Age Group × Consistent Versus Inconsistent Outcome × Causal Explanation Type interaction was significant, $F(4, 84) = 2.74$, $p < .05$. In post hoc simple effects analyses of children's explanations for inconsistent outcomes only, as shown in Figure 5, causal function ($M = 1.44$ out of 4 trials or 36%) and causal action ($M = 0.23$ out of 4 trials or 6%) explanations remained constant across age groups. Category switch explanations, however, increased significantly with age (16%, 38%, and 53% of total explanations at ages 3, 4, and 5, respectively), $F(2, 47) = 3.95$, $p < .05$. The percentage of participants that gave category switch, causal function, and causal action explanations at least once was calculated. These data are displayed in Table 3a. This analysis is a useful supplement to the data on proportion of total explanations of each explanatory type, because it demonstrates that a substantial portion of 3- and 4-year-olds make use of this kind of explanation. This finding is especially noteworthy given that category switch explanations require overriding perceptual appearances and prior knowledge about an object and spontaneously redefining category boundaries around function.

For consistent outcomes (not shown in Figure 5), whereas causal function explanations ($M = 1.19$ out

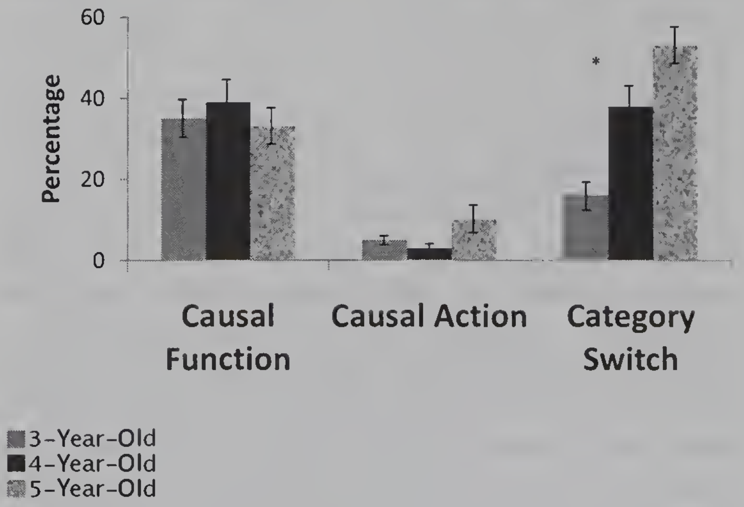


Figure 5. Study 1: Percentage of explanation types for inconsistent outcomes by age group.
Note. Comparison between age groups.
$^{*}p < .05$.

Table 3a

Study 1: Percentage of Children Giving Each Kind of Causal Explanation for Inconsistent Outcomes at Least Once by Age Group

	Category switch	Causal function	Causal action
3-year-olds	31%	56%	19%
4-year-olds	56%	69%	6%
5-year-olds	69%	56%	19%

Table 3b

Study 2: Percentage of Children Giving Each Kind of Causal Explanation for Inconsistent Outcomes at Least Once by Age Group

	Category switch	Causal function	Causal action
3-year-olds	25%	94%	13%
4-year-olds	69%	50%	19%

of 4 trials or 30%) and causal action explanations ($M = 0.21$ or 5%) remained constant across age groups, $F(2, 47) = 1.47$, *ns*; $F(2, 47) = 1.78$, *ns*, category label explanations increased with age (6%, 24%, and 31% of total explanations at ages 3, 4, and 5), $F(2, 47) = 3.56$, $p < .05$. The increase in category label explanations with age for consistent outcomes parallels the increase in category switch explanations with age for inconsistent outcomes.

Discussion

The test trial data from Study 1 demonstrate that inconsistent outcomes are an especially powerful trigger for children's explanations. Given that children had readily available explanations for the consistent outcome, they could have easily ignored the inconsistent outcome and just explained the consistent outcome. Instead, children found outcomes inconsistent with prior knowledge especially noteworthy and were much more likely to explain an inconsistent outcome first, as well as more likely to provide multiple explanations for an inconsistent outcome overall. This pattern held across the generative and inhibitory conditions. Further, if children had just been paying attention to the perceptual outcome (such as state change or negative outcome), the event they explained should have been at chance, given that in the test trials the outcomes were identical (both boxes were on or off). Indeed, data from the confirmation trials showed that in the absence of an inconsistent event or outcome, state change did serve as a trigger for children's explanations. But on test trials children explained inconsistent outcomes whether they involved state change or not. Finally, one could argue that children were primed to provide an explanation for the *consistent* outcomes prior to the test trials, because the confirmation trial (which appeared first) asked children to explain outcomes that were consistent with their prior knowledge about how the objects function. Nonetheless on test trials children systematically explained inconsistent events.

At the same time, it was not the case that children never explained consistent outcomes and therefore that children's explanations were never confirmatory. In their overall explanations, children did competently and easily provide explanations for consistent events as well. Constructing explanations for events that are consistent with children's prior knowledge and experience must at times also be important, providing children with opportunities to deepen their understanding of expected phenomena. (Indeed, in our data children often did generate "deeper" explanations of consistent events by generating causal function explanations that cited possible mechanisms, such as underlying causal properties and internal components.) In short, we propose that outcomes inconsistent with prior knowledge are especially powerful triggers for causal explanatory cognition, not that children do not or cannot generate explanations for outcomes consistent with their prior knowledge.

Exactly how children explain events provides further insight into their understanding of the causal basis for the outcome. Children provided three distinct kinds of causal explanations: causal function, category switch (or category label in the case of consistent outcomes), and causal action explanations. Across age groups, causal function explanations were given quite frequently overall, with more than half of the children at each age providing at least one such explanation. Causal function explanations referred to alternative or problematic functioning for inconsistent outcomes and referred to the object functioning correctly for consistent outcomes. Causal function explanations often included information about unobserved inside parts (power, batteries) or underlying causal properties (energy, electricity). For example, a 3-year-old explained, "Because this ran out of lights. Because it doesn't have light stoppers in it," and according to a 4-year-old, "The starter isn't working right, its inside parts must be all broken."

Category switch explanations for inconsistent outcomes were also frequent (e.g., "It's really a starter; it only looks like a do-nothing" or "Hey! This one is a starter too! I thought it wasn't but it is!"). These explanations are noteworthy because they entail spontaneously redefining category boundaries and privileging functional information over appearance. When explaining consistent outcomes, children also evoked the category label (e.g., "Because it's a starter. Starters make the light in the boxes turn on"), but without these additional features. Category switch explanations increased with age, and by 5 years of age, more than two thirds of participants provided at least one such explanation for inconsistent outcomes.

Prior research shows that even 3-year-olds can categorize objects in terms of novel, nonobvious properties (Gelman & Coley, 1990; Gelman & Markman, 1986, 1987), can apply names to objects with the same functional properties (Kemler-Nelson, 1995), and can categorize and name objects based on novel causal properties (Gopnik & Sobel, 2000). What is unique about the present task is that children spontaneously generated nonobvious, invisible properties for items and recruited this sort of reasoning to explain why items functioned in both expected and unexpected ways.

Several open questions remain, in particular about our use of functional labels in Study 1. Children's responses might have been scaffolded by the object labels (starters, stoppers, and do-nothings), which were straightforward descriptions of their functions. For example, during inconsistent events, rather than truly noting the inconsistent outcomes, children instead may have noted the inconsistency of the observed function compared to the (inferred) label. Additionally, among the most interesting kinds of explanations provided by children in Study 1 were category switch explanations. However, given the nature of the functional label, it is unclear whether children were actually relabeling on the basis of causal effects or whether instead they were simply giving a more literal description of function (i.e., "it's not really a starter" may be equivalent to "it didn't start it"). Study 2 was designed specifically to address these possibilities.

Study 2

Study 2 addressed whether explanations for events not in accord with prior knowledge were scaffolded by the use of functional labels (*starters*, *stoppers*, and *do-nothings*) in Study 1, because the objects violated labeled functions. To investigate this, in Study 2 we used novel, nonfunctional labels for the objects (*toma*, *blicket*). Given that even 2-year-olds can extend novel labels based on function (Kemler-Nelson, 1995), we expected that children in Study 2 would also make use of information about function and underlying causal properties, much as in Study 1, even when the labels are novel and do not, in themselves, provide additional causal information. Additionally, based on Study 1, we predicted that children would again focus more on explaining events inconsistent with prior knowledge, even when the labels themselves do not carry any functional information.

Method

Participants

Given the near ceiling performance by 5-year-olds on several measures in Study 1, here we focused on younger children. Sixteen 3-year-olds (*M* age = 3.5; range = 3.0–3.10) and sixteen 4-year-olds (*M* age = 4.4; range = 4.0–4.11) who had not participated in Study 1 were recruited from a Midwestern U.S. university town. Participants were primarily White and from middle-class families. Approximately equal numbers of boys and girls participated in the study.

Materials

The materials used in Study 2 were identical to those from Study 1.

Procedure

The procedure for Study 2 was modeled on Study 1, with a single critical modification: The objects were given nonfunctional, novel labels (*a toma/not a toma* and *a blicket/not a blicket*), and were never referred to as *a starter*, *a stopper*, or *a do-nothing*. The confirmation and test trials were identical to those of Study 1.

Transcribing and Coding

The same procedure for coding Study 1 was amended slightly for Study 2 (to encompass reference to tomas and blickets instead of starters and stoppers). Table 2 provides examples of category switch, causal function, and causal action explanations from Study 2. Interrater reliability was established using a randomly selected sample of 25% of the data. Two persons independently coded the reliability sample with 98% agreement. Reliability was calculated for explanatory coding categories for both consistent and inconsistent outcomes, with kappas ranging from .92 to .94, all within near perfect levels (Landis & Koch, 1977).

Results

As in Study 1, we assessed whether events inconsistent with prior knowledge led to greater interest and attention, and most importantly, were first and most often explained.

Which Outcomes Do Children Look at and Explain First?

Confirmation trials. In the confirmation trials, children were more likely to look first at the object that changed the state of the box (started or stopped it) (64% of the time), $t(31) = 2.06$, $p < .05$, and to first explain the object that changed the state of the box (62% of the time), $t(31) = 1.97$, $p < .05$, when compared with the object that had no effect on the box. Although children were more likely to first explain the object that changed the state of the box in both conditions, in the generative condition ($M = 70\%$) they did so above chance, $t(31) = 2.63$, $p < .01$, but in the inhibitory condition ($M = 53\%$) they were at chance, $t(31) = 0.32$, *ns*. One possible explanation is that children are more compelled by an object that fails to turn the light box off than an object that fails to turn the light box on. More importantly, and as in Study 1, in the generative condition, where the confirmation trial pitted a state change event against a negative outcome, children were significantly more likely to first explain the state change outcome than the negative outcome.

Test trials. Again, inconsistent outcomes both attracted children's attention first and provoked children's explanations. Across age groups, children were much more likely to look first at the inconsistent outcomes than the consistent outcomes ($M = 3.25$ out of 4 or 81%), $t(31) = 7.44$, $p < .001$. Likewise, children were much more likely to explain the inconsistent outcome first ($M = 3.05$ or 76%), $t(31) = 9.86$, $p < .001$.

In order to investigate potential age- and condition-related effects, an Age Group (3- and 4-year-olds) × Condition (inhibitory, generative) × Final Outcome (both boxes on, both boxes off) repeated measures ANOVA was conducted. Condition and final outcome were within-subjects factors and age group was a between-subjects factor. As in Study 1, at all age groups and conditions, children were very likely to look at (3-year-olds 83%, 4-year-olds 80%) and explain (3-year-olds 75%, 4-year-olds 78%) the inconsistent outcome first. Thus, there was no significant main effect of age (for either the first look or first explanation analysis). There were also no significant main effects of condition (inhibitory vs. generative) or final outcome (both on vs. both off). Moreover, unlike Study 1, none of the interactions were significant.

As in Study 1, to investigate the possibility that the sequential structure of the task may have influenced the outcome children were motivated to explain first, across conditions, the outcome children explained in the test trials they were presented with first (following the confirmation trial) were compared to the outcome they explained in the test trials they were presented with second. There were no significant differences for test trial order, $t(31) = .71$, *ns*.

Spontaneous explanations children provided prior to an explicit explanatory question were also analyzed as a supplementary analysis. As in Study 1, spontaneous explanations revealed the same pattern of responses as elicited explanations for first explanation. Ten of the 32 children participating provided a spontaneous explanation at least once in the test trials (31%). For 90% of the spontaneous explanations children explained the inconsistent outcome first, $F(1, 31) = 185.61$, $p < .001$. Children's interest in explaining inconsistent outcomes even without an explicit explanatory question provides clear evidence that children were not merely explaining what they thought the experimenter was interested in.

Which Outcomes Do Children Provide an Explanation for Overall?

As in Study 1, children often provided multiple explanations. When focused on the total number of explanations provided, children were again much more likely to explain events inconsistent with their prior knowledge than events consistent with their prior knowledge. In order to investigate potential age and condition-related effects, we conducted a 2 (age group: 3- and 4-year-olds) × 2 (condition: inhibitory, generative) × 2 (final outcome: both boxes on, both boxes off) × 2 (consistent vs. inconsistent outcome) repeated measures ANOVA on the entirety of children's explanations. Condition, final outcome, and consistent versus inconsistent outcome were within-subjects factors. The only significant effect was for consistent versus inconsistent outcome, $F(1, 30) = 22.64$, $p < .001$, reflecting that inconsistent outcomes were more likely to be explained than consistent outcomes ($M = 3.94$ or 99% for inconsistent outcomes and $M = 2.78$ or 70% for consistent outcomes). As shown in Figure 6, both 3-year-olds, $t(15) = 3.47$, $p < .01$, and 4-year-olds, $t(15) = 2.7$, $p < .05$, showed this effect. The

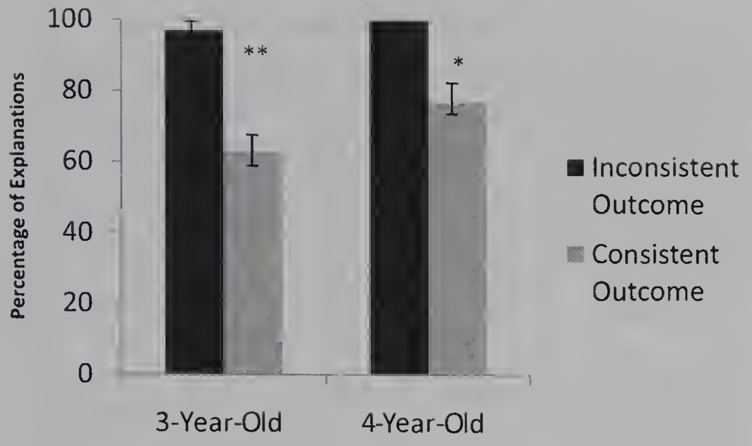


Figure 6. Study 2: Percentage of explanations children provided for inconsistent versus consistent outcomes by age group.
Note. Comparisons between explanations for consistent versus inconsistent outcomes.
$^{*}p < .05$. $^{**}p < .01$.

lack of any significant effects for age group and condition may reflect the more narrow age range in Study 2 (only 3- and 4-year-olds) in comparison with Study 1 (3-, 4-, and 5-year-olds).

The Kinds of Causal Explanations Children Provided

As in Study 1, the majority of children's explanations at all age groups were causal, for both inconsistent outcomes (81% and 84% at ages 3 and 4, respectively) and consistent outcomes (63% and 74% at ages 3 and 4). An Age Group (3- and 4-year-olds) × Condition (inhibitory, generative) × Consistent Versus Inconsistent Outcome × Causal Explanation Type (causal function, causal action, category label/category switch) repeated measures ANOVA was conducted in order to investigate age-related changes in the kinds of causal explanations children provided. Condition, consistent versus inconsistent outcome, and causal explanation type were within-subject factors. The number of total explanations provided (of each type) was the dependent variable.

As predicted, and as noted earlier, there was a significant main effect for consistent versus inconsistent outcome, $F(1, 30) = 25.94$, $p < .001$. The main effect for causal explanation type was also significant, $F(2, 60) = 9.77$, $p < .001$. Post hoc tests indicate that children were more likely to provide causal function explanations than causal action explanations, *ps* < .01. This can be seen in Figure 7.

The three-way Age Group × Consistent Versus Inconsistent Outcome × Causal Explanation Type interaction was also significant, $F(2, 60) = 5.64$, $p < .01$. This interaction indicates, as in Study 1, that there were changes with age in the kinds of causal explanations that children provided for inconsistent outcomes. Whereas causal action explanations remained constant (and infrequent) across age groups, causal function explanations decreased with age, $p < .001$, and category switch explanations increased with age, $p < .01$ (see Figure 7). We also examined the number of participants that gave category switch, causal function, and causal action explanations for inconsistent outcomes at least once. These results are informative because, as in Study 1, they indicate that a substantial portion of 3-year-olds and most 4-year-olds provided at least one category switch explanation, but also that 3-year-olds were very likely to provide explanations based on the causal functioning of the objects, even though the objects were given novel, *nonfunctional* labels (Table 3b).

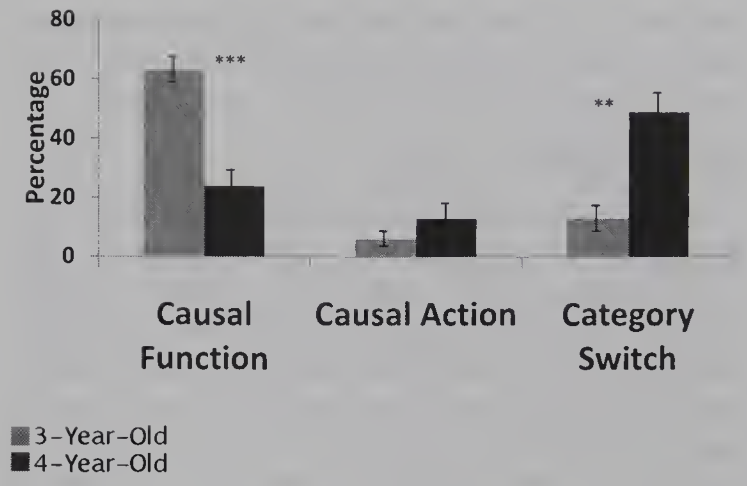


Figure 7. Study 2: Percentage of explanations types for inconsistent outcome by age group.
Note. Comparisons between age groups.
$^{**}p < .01$. $^{***}p < .001$.

For consistent outcomes (not shown in Figure 5), whereas causal function explanations and causal action explanations remained constant across age groups, category label explanations increased with age, $p < .05$. As in Study 1, this parallels the increase in category switch explanations with age for inconsistent outcomes.

Additional post hoc tests revealed that 3-year-olds were more likely to give causal function explanations for inconsistent ($M = 2.5$ out of 4 or 63%) than consistent outcomes ($M = 1.4$ or 35%), $p < .01$. Four-year-olds were more likely to give causal action explanations for inconsistent (13%) than consistent outcomes (3%), $p < .05$, and more likely to give category switch explanations for inconsistent outcomes (48%) than category label explanations for the consistent outcomes (24%), $p < .01$. Explanations that made use of psychological causes/descriptions were very rare; only two 3-year-olds and one 4-year-old used psychological language in their explanations (''This one doesn't know how to act; it isn't working right'').

Comparing Studies 1 and 2

Overall, the results from Studies 1 and 2 were highly similar. Across both studies, children's explanations on the confirmation trials focused on state changes. For test trials for both studies, inconsistent outcomes were especially likely to provoke children's explanations. Moreover in both studies, the content of children's explanations, specifically category and causal function explanations, consistently referred to underlying, internal causal mechanisms, overriding perceptual appearances. This is especially noteworthy for Study 2, where the labels were novel and provided no explicit functional information. Thus, the results of Study 1 cannot be attributed merely to the use of functional labels. One difference did occur: Children were more likely to look at the inconsistent outcome first in Study 1 (93%) than in Study 2 (81%), $F(1, 79) = 7.24$, $p < .01$, indicating that the functional labels seem to have increased children's attention to the inconsistent outcome. However, in both studies, children were equally likely to explain the inconsistent outcome first (80% in Study 1 and 76% in Study 2). Thus, children's explanations of inconsistent outcomes depend on noticing those outcomes, but language can influence children's attention and explanation.

Discussion

Study 2 further demonstrates that children focus more on explaining events inconsistent with prior knowledge than events consistent with prior knowledge, even when the labels themselves do not provide functional information. Moreover, children's causal explanations make reference to function and underlying causal properties. As in Study 1, children privileged function information over perceptual features; evoked causal mechanisms and unobservable, internal causal properties; and spontaneously used (novel) labels to redefine category boundaries around shared function. Thus, several key results of Study 1 were not merely redescription based on the functional labels provided there. Finally, Study 2 provides evidence that use of functional labels was not what drew children to explain inconsistent events (because of the inconsistency of the outcome with a functional label), but instead children are more generally motivated to provide explanations for outcomes that were inconsistent with their prior knowledge.

General Discussion

Causal explanation is a goal-directed cognitive activity that depends on what is relevant or important to the explainer. Given the potentially unlimited number of outcomes one could explain, especially for young children early in the process of developing causal knowledge, identifying the specific kinds of events, outcomes, and goals that trigger explanation provides insight into what motivates and guides causal cognition. Although a substantial amount of developmental literature has demonstrated that young children frequently seek (Callanan & Oakes, 1992) and provide explanations (Wellman et al., 1997), little is known about what triggers this childhood explanatory activity.

If children have a cognitive model of the world based on a framework of anticipatory regularities, they would be well equipped to rapidly form expectations contingent on prior beliefs or knowledge. Indeed, even infants differentially attend to and explore events or objects inconsistent with such anticipatory regularities (Baldwin, Markman, & Melartin, 1993). In our data so do preschoolers; they first attend to inconsistent over consistent outcomes. Attention is not explanation, however. Research on confirmation bias in scientific reasoning provides evidence that patterns of attention do not guarantee sufficient motivation to generate explanations or promote causal learning (Dunbar & Klahr, 1988; Kuhn, 1989). Children could easily note inconsistent outcomes but be motivated primarily to explain consistent outcomes. Because such events conform to known expectations, they provide a foundation for constructing more generalizable causal explanations—particularly for young learners. In contrast, our data demonstrate that inconsistent outcomes were an especially powerful trigger for young children's explanations. Across both generative and inhibitory conditions, with meaningful functional labels and with arbitrary novel labels, children were much more likely not only to attend to but also to explain the inconsistent outcome.

In the course of this research we considered additional factors possibly relevant to children's explanations (such as state change or negative outcomes). An important feature of the test trials in Studies 1 and 2 was that they controlled for these additional potential explanatory triggers by holding them constant. Confirmation trials, however, allowed examination of state change or negative outcome as a potential explanatory trigger in the absence of an inconsistent outcome. Although inconsistency with prior knowledge is a compelling trigger for children's causal explanations, in the absence of this information, state change was an additional explanatory trigger. We suggest that attention to state change may hold generally, beyond the particulars of this experiment. For example, children may expect individuals to stay healthy or objects to remain in motion—in which case state changes, such as contracting illness, may trigger children's explanations.

Another interesting possibility is that there may be differences in what one attends to, depending on where one is in the theory construction process. It is quite striking that children attend to and explain inconsistent outcomes after only minimal exposure to the contingencies. It may be the case that earlier in the theory construction process children are more receptive to inconsistent information and less so later on. Schema research supports this possibility. The memory advantage of schema-inconsistent items became weaker as the proportion of anomalous or schema-inconsistent items becomes larger (Rojahn & Pettigrew, 1992). Attending to frequency information would also help reconcile the motivation to seek causal regularities while responding to inconsistent information. Research on clustering accounts of causal learning predicts that inconsistent items are better remembered because they are differentiated from clusters that capture regularities (Sakamoto & Love, 2004). Frequency information may then drive the decision to integrate inconsistent information into an existing cluster or schema or create an additional schema to accommodate the new information. Further exploration of the conditions in which children focus on consistent information versus inconsistent information would be an interesting direction for future research.

More generally, an important open question is the role that explanatory processes play in children's development. There is mounting evidence that children's causal explanations may in fact constitute a mechanism for causal learning (Chi et al., 1989). When children are asked to explain events, they learn better than when given simple feedback about whether their predictions are correct (e.g., Amsterlaw & Wellman, 2006; Chi, DeLeeuw, Chiu, & LaVancher, 1994; Siegler, 1995). But how does this work? If explanation is a mechanism for learning, children should provide explanations for events that have the potential to teach them something new. Events that are inconsistent with prior knowledge provide just such an opportunity. Such an explanatory tendency could aid in learning by focusing children on events that challenge current causal knowledge and provoke additional, amended causal reasoning. Given that mere attention to inconsistent information is insufficient to ensure causal learning or theory revision (Kuhn, 1989), explanation may serve as a critical mechanism for integrating and reconciling inconsistent information with existing theories or knowledge structures.

The nature of children's explanations, in appealing to unobserved explanatory processes and functional properties, could similarly aid learning. Explanations that appeal to underlying causes and mechanisms engage children in the important interplay between data and theory that leads to theory change (Wellman & Gelman, 1998). Children's explanations for inconsistent outcomes, in particular, systematically referred to such features. For example, as shown in Table 2, children explained inconsistent outcomes by stating that "All the energy is gone" or "Maybe that toma doesn't have batteries in it." Additionally, category switch explanations (e.g., "It looks like a toma, but it's not. It's not really a toma at all") indicate that even 3-year-olds privilege functional information over appearances.

Our data offer insight into the function of children's explanations by providing empirical evidence for the kinds of events that guide causal cognition by motivating children to construct explanations. Although the present experiments do not directly test the hypothesis that children's explanations play an active role in the learning process, they are consistent with this proposal and provide a promising empirical basis for investigating the interplay of explanation and learning in ongoing and future research.

References

Amsterlaw, J., & Wellman, H. (2006). Theories of mind in transition: A microgenetic study of the development of false belief understanding. *Journal of Cognition and Development*, *7*, 139–172.

Baillargeon, R. (2002). The acquisition of physical knowledge in infancy: A summary in eight lessons. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (pp. 47–83). Oxford, UK: Blackwell.

Baldwin, D., Markman, E., & Melartin, R. (1993). Infants' ability to draw inferences about nonobvious object properties: Evidence from exploratory play. *Child Development*, *64*, 711–728.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. Oxford, UK: Wiley.

Callanan, M., & Oakes, L. (1992). Preschoolers' questions and parents' explanations: Causal thinking in everyday activity. *Cognitive Development*, *7*, 213–233.

Chi, M., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13*, 145–182.

Chi, M., DeLeeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*, 439–477.

Dunbar, K., & Klahr, D. (1988). Developmental differences in scientific discovery strategies. In D. Klahr & K. Kotovsky (Eds.). *Complex information processing: The impact of Herbert A. Simon. Proceedings of the 21st Carnegie-Mellon Symposium on Cognition* (pp. 109–144). Hillsdale, NJ: Erlbaum

Gelman, S. A., & Coley, J. (1990). The importance of knowing a dodo is a bird: Categories and inferences in 2-year-old children. *Developmental Psychology*, *26*, 796–804.

Gelman, S. A., & Markman, E. (1986). Categories and induction in young children. *Cognition*, *23*, 183–209.

Gelman, S. A., & Markman, E. (1987). Young children's inductions from natural kinds: The role of categories and appearances. *Child Development*, *58*, 1532–1541.

Gopnik, A., & Sobel, D. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. *Child Development*, *71*, 1205–1222.

Gopnik, A., Sobel, D., Schulz, L., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, *37*, 620–629.

Graham, S. A., Kilbreath, C. S., & Welder, A. N. (2004). Thirteen-month-olds rely on shared labels and shape similarity for inductive inferences. *Child Development*, *75*, 409–427.

Hempel, C. (1965). *Aspects of scientific explanation*. New York: Free Press.

Hickling, A., & Wellman, H. (2001). The emergence of children's causal explanations and theories: Evidence from everyday conversation. *Developmental Psychology*, *37*, 668–683.

Hilton, D. (1995). Logic and language in causal explanation. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 495–529). Oxford, UK: Clarendon Press.

Jaswal, V. K., & Markman, E. M. (2007). Looks aren't everything: 24-month-olds' willingness to accept unexpected labels. *Journal of Cognition and Development*, *8*, 93–111.

Keil, F. (2006). Explanation and understanding. *Annual Review of Psychology*, *57*, 227–254.

Kemler-Nelson, D. (1995). Principle-based inferences in young children's categorization: Revisiting the impact of function on the naming of artifacts. *Cognitive Development*, *10*, 347–380.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, *96*, 674–689.

Kushnir, T., & Gopnik, A. (2005). Young children infer causal strength from probabilities and interventions. *Psychological Science*, *16*, 678–683.

Kushnir, T., & Gopnik, A. (2007). Conditional probability versus spatial contiguity in causal learning: Preschoolers use new contingency evidence to overcome prior spatial assumptions. *Developmental Psychology*, *43*, 186–196.

Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174.

Legare, C. H., Wellman, H. M., & Gelman, S. A. (2009). Evidence for an explanation advantage in naïve biological reasoning. *Cognitive Psychology*, *58*, 177–194.

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Science*, *10*, 464–470.

Medin, D., Coley, J., & Storms, G. (2003). A relevance theory of induction. *Psychonomic Bulletin & Review*, *10*, 517–532.

Mills, C., & Keil, F. (2004). Knowing the limits of one's understanding: The development of an awareness of an illusion of explanatory depth. *Journal of Experimental Child Psychology*, *87*, 1–32.

Nickerson, R. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*, 175–220.

Popper, K. R. (1963). *Conjectures and refutations: The growth of scientific knowledge*. New York: Harper & Row.

Rojahn, K., & Pettigrew, T. (1992). Memory for schema-relevant information: A meta-analytic resolution. *British Journal of Social Psychology*, *31*, 81–109.

Sakamoto, Y., & Love, B. (2004). Schematic influences on category learning and recognition memory. *Journal of Experimental Psychology: General*, *133*, 534–553.

Schulz, L., & Gopnik, A. (2004). Causal learning across domains. *Developmental Psychology*, *40*, 162–176.

Shultz, T. R. (1982). Rules of causal attribution. *Monographs of the Society for Research in Child Development*, *47*, (Serial No. 194).

Siegler, R. (1995). How does change occur: A microgenetic study of number conservation. *Cognitive Psychology*, *28*, 225–273.

Sloman, S. A. (1994). When explanations compete: The role of explanatory coherence on judgments of likelihood. *Cognition*, *52*, 1–21.

Weiner, B. (1985). ''Spontaneous'' causal thinking. *Psychological Bulletin*, *97*, 74–84.

Wellman, H. M., & Gelman, S. A. (1998). Knowledge acquisition in foundational domains. In D. Kuhn & R. Siegler (Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (5th ed., pp. 523–573). New York: Wiley.

Wellman, H., Hickling, A., & Schult, C. (1997). Young children's psychological, physical, and biological explanations. In H. M. Wellman & K. Inagaki (Eds.), *The emergence of core domains of thought: Reasoning about physical, psychological and biological phenomena*, (pp. 7–25). San Francisco: Jossey-Bass.

Wilson, R., & Keil, F. (1998). The shadows and shallows of explanation. *Minds and Machines*, *8*, 137–159.

Wong, P., & Weiner, B. (1981). When people ask ''why'' questions, and the heuristics of attributional search. *Journal of Personality and Social Psychology*, *40*, 650–663.

Child Development, May/June 2010, Volume 81, Number 3, Pages 945–957

Children's Tattling: The Reporting of Everyday Norm Violations in Preschool Settings

Gordon P. D. Ingram and Jesse M. Bering
Queen's University Belfast

Tattling, defined as the reporting to a second party of norm violations committed by a third party, is a frequent but little-studied activity among young children. Participant observation and quantitative sampling are used to provide a detailed characterization of tattling in 2 preschools (initial mean age = 4.08 years, $N = 40$). In these populations, tattling represents the majority of talk about peers' behavior to third parties. It is usually truthful, it rarely refers to transgressions committed against other individuals, it is not often ignored by adults, it is performed more frequently by dominant children, and it correlates with teacher reports of relational aggression. These exploratory results suggest several new avenues of research into children's developing understanding of social norms.

Moral understanding does not develop in isolation from other competences, but is used by children to navigate everyday social interactions (for overviews, see Dunn, 2006; Turiel, 2006). An example is the use of normative appeals (e.g., "That's not fair!") in attempts to resolve social conflicts. Naturalistic observation has proved useful in analyzing such behavior, dating back to Piaget (1932), who studied appeals to social norms in the context of children's street games. In family contexts, Dunn (1988) carried out extensive naturalistic studies of young children's spontaneous justification and condemnation of everyday social behavior (see especially Dunn & Munn, 1987). Despite Piaget's lead, such activities have received less attention in peer contexts, although Nucci and Turiel (1978) did observe the responses of preschool children to moral and conventional transgressions by peers in naturalistic settings. One of the categories of responses that they considered was that of verbal reports to the teacher. This kind of response is interesting because it represents the use of evaluative language to recruit third-party involvement in social problems, suggesting an early awareness of the moral force of language.

In the present article, we link preschool children's everyday language use to their social and moral development through a naturalistic study of *tattling*. We define tattling as *the reporting to a second party of a third party's counternormative behavior*, where "counternormative" may refer to explicitly proscribed behavior, or to behavior that is unwelcome to the individual who reports it and that they implicitly believe will also be unwelcome to their audience. The structure of tattling therefore offers a window on how individual children start to relate to social norms and to incorporate them in everyday discourse.

Social norms can operate effectively only if they are taught to new generations of children. Nichols (2004) has argued that norms are more likely to be selected for—through cultural selection—if they can effectively pass through the "cognitive filter" of children's learning processes, which are more likely to happen if they produce a strong affective response. Independent of this affective component, Nichols also postulated that children are primed to acquire a "normative theory" early in development, suggesting that even very young children are sensitive to normative statements by adults. He cited the experiments of Harris and Núñez (1996), who showed that 3- to 4-year-old children were especially sensitive to actions that breached a

---

This research was supported by a grant from the Northern Ireland Department of Employment and Learning, as part of the first author's doctoral dissertation. We thank the teachers, classroom assistants, and children for participating in the study; Bethany Heywood and Jared Piazza for coding the observational data; Nicola Ingram and Paulo Sousa for reading and commenting on the manuscript; and three anonymous reviewers at *Child Development* for their many helpful suggestions.

Correspondence concerning this article should be addressed to Gordon P. D. Ingram, Institute of Cognition and Culture, Queen's University Belfast, Belfast BT7 1NN, UK. Electronic mail may be sent to gingram02@qub.ac.uk.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8103-0020

permission rule, compared to actions that breached a description rule. Their participants were better at identifying a picture where a child was "doing something naughty" than a picture where a child was "doing something different" (1996, Experiment 4; cf. Cosmides, 1989). Results similar to this have been found in cross-cultural studies with children (Harris, Núñez, & Brett, 2001) and adults (Sugiyama, Tooby, & Cosmides, 2002), suggesting that there may be an innate cognitive bias toward identifying breaches of social norms. Harris (2000) has pointed out that young children seem to generalize very readily about normative obligations: "Any novel obligation is understood in the light of a well-organized, pre-existing concept of constraint" (p. 158). The same point was made by Rakoczy, Warneken, and Tomasello (2008), who conducted experiments in which an experimenter demonstrated novel games (e.g., "daxing"), played by unfamiliar rules, to 2- and 3-year-old children. When a puppet started playing and broke these rules, the 3-year-olds spontaneously made both "normative protests" (e.g., "That's not how you do it!") and "imperative protests" (e.g., "Don't do that"). The 2-year-olds were apparently too young to make a significant number of normative protests, but even at this early age they were given to making imperative protests about the puppet's behavior.

If young children are especially sensitive to norm violations in an experimental context, the same cognitive bias might influence their natural verbal behavior. We suspected that the general sensitivity toward norm violations revealed by these experiments would make young children more likely to report everyday norm violations by peers. This led us to predict that tattling (as broadly defined above) would make up a large part of children's everyday communication about third-party behavior, and that a study of the general characteristics of tattling would reveal interesting features of children's attitudes to social norms.

Although analogous activities—such as snitching (Rosenfeld, Jacobs, & Wright, 2003), whistleblowing (Brewer & Selden, 1998), or eyewitness testimony (Loftus, 1979)—are practiced by adults when they report criminal behavior to an authority figure, the word *tattling* is usually associated with young children's behavior. For much of the 20th century it was disapproved of by educators, and seen as a practice that children needed to be educated away from (Williams, 1989). While not infrequently mentioned in studies of child development, tattling has rarely been analyzed in depth, but more often simply listed as an example of undesirable behavior (e.g., Hurlock & McDonald, 1934; McConnell, 1963). A more detailed early treatment was by Smith (1932), who noted that negative "criticism" of a third party by children was much more common than positive comments, and that unlike adults, children often made their criticisms within earshot of the target. This suggests that children did not see anything counternormative about the act of tattling, even in an era when tattling was likely to be frowned upon by adults.

Apart from Smith's (1932) early and rather unsystematic study of children's criticisms, no previous observational study seems to have focused on tattling in an educational setting. The only modern observational study devoted to tattling was by Ross and den Bak-Lammers, who carried out a longitudinal study of sibling dyads aged 2 and 4 (Den Bak & Ross, 1996), and 4 and 6 (Ross & den Bak-Lammers, 1998). Siblings were observed in their family homes, and all their utterances recorded for later analysis. For all the age groups studied, tattling made up a large proportion of talk about the sibling's behavior. This proportion decreased with age (although tattling did increase in absolute frequency), ranging from 87.1% among the 2-year-olds, through 74.8% among the 4-year-olds, to 56.4% among the 6-year-olds. One of the main aims of the present study was to find out whether the preponderance of tattling in young children's communication about siblings would generalize to their communication about unrelated peers.

Tattling on peers is quite prevalent among children in elementary school settings, as Skinner, Cashwell, and Skinner (2000) anecdotally attested for the 9- to 10-year-old children with whom they studied. Skinner and his colleagues had some success with their program to encourage "tootling" (a term they coined for the reporting of prosocial behavior) by providing a collective reward for the class when they reached a target number of "tootles" (see also Morrison & Jones, 2007). Implicit in their account was the point that tattling, unlike tootling, does not have to be extrinsically rewarded in order to take place but presumably has some sort of intrinsic motivation. In some situations tattling may be motivated by processes of emotion regulation (Thompson, 1994). Cooney, Hutchison, and Costigan (1996) postulated that toddlers' tattling is an intermediate stage in emotion regulation, which supersedes direct physical aggression and precedes more sophisticated forms of negotiation with peers that do not rely on adult intervention.

By adolescence, the overt reporting of peers' transgressions to adult authority figures seems to become much less common. In a questionnaire-based study of tattling among teenagers in a residential care program, Friman et al. (2004) found that perceived rates of tattling correlated negatively with likeability and positively with social rejection. On the other hand, evaluative talk about peers to other peers becomes more important during preadolescence and adolescence (e.g., Cameron, 1997; Fine, 1986; Goodwin, 1990). Several groups of researchers have considered evaluative talk among these age groups in terms of the theoretical constructs of *relational*, *indirect* and *social aggression* (reviewed by Archer & Coyne, 2005). Central to the first two of these constructs, and often implicit in the third, is the idea that such verbal aggression is often covert, so that perpetrators are able to conceal their identities from their victims.

However, few observers have recorded covert talk about others' behavior among children younger than 9 years. In a discussion of children's gossip, Fine (1977) pointed out that early forms of talk about peers are rarely covert: "One salient difference between the social structure of adult gossip and that of children is that adult gossip is virtually always about non-present others, whereas children often gossip in front of the target" (p. 183). Mettetal (1983) conducted quantitative, naturalistic research into children's everyday talk about peers, and found that the frequency of gossip-like discussions between dyads of girls increased dramatically between the ages of 6–7 and 11–12, rising to about one third of all conversations. A similar pattern was found by Engel and Li (2004), who asked three groups of children—aged 4, 7, and 10—to tell stories about their friends in semistructured interviews. The length, descriptiveness, and evaluative content of the stories all increased significantly with age, implying that the younger children's stories were far less informative than the older children's. This supported Engel and Li's naturalistic observation, from tape recordings of conversations in a day-care center, that 4-year-old children very rarely told stories about absent peers: "It was surprisingly difficult to catch the children gossiping" (p. 160). Coupled with the observation of Ross and den Bak-Lammers (1998) that tattling occupies a decreasing proportion of overall communication about sibling behavior, these results suggest that overt tattling may decrease in importance through middle childhood at about the same time that covert gossip increases.

From this brief survey of the existing literature on tattling and other forms of evaluative social discourse in children, the following points informed our research design. First, tattling is a widespread activity among young children, but one that has been little studied in any context—and that has never been systematically studied in the context of the preschool classroom. Second, as children grow older, tattling to adults becomes both less frequent and less overt: It tends to have negative reputational consequences, whereas the covert reporting of peers' activities to other peers becomes more common. Third, tattling may involve the socialization of aggressive impulses, representing an intermediate stage between physical violence and negotiation when a child is confronted with unwelcome behavior from a peer. With these points in mind, we carried out an exploratory observational study of tattling in the preschool classroom. As tattling had never been systematically studied in the preschool context, our study sought principally to establish a baseline of data about the general characteristics of young children's tattling within peer groups. The aim was to compare the results with existing data on tattling in the home, and ultimately to generate hypotheses for more focused observational and experimental work.

Method

We used a mixture of quantitative and qualitative empirical approaches (Tashakkori & Teddlie, 2003) to characterize as many different properties of tattling as possible. The first author engaged in participant observation of the children while working as a volunteer classroom assistant in two preschools. For about half of the observation time, he took on the more detached role of an independent observer and conducted behavioral sampling. The qualitative and quantitative methodologies were complementary in that the former helped us understand the meanings and motivations of acts of tattling for both children and adults, whereas the latter allowed us to quantify various descriptive properties of tattling.

Participants

Research was conducted in two inner-city preschools in Belfast, the principal city of Northern Ireland. The first preschool studied (Preschool A) was situated in a low-income residential area of Belfast, inhabited overwhelmingly by working-class Catholics. All the children came from (Northern) Irish families, with the exception of one child who was of mixed Irish/Portuguese parentage. When research began in November 2006, there were 15 children in the preschool (8 boys and 7 girls). Their age at the start of the research, which lasted for three calendar months, ranged from 3.52 to 4.32 years ($M = 3.97$). In January 2007 they were joined by a younger girl aged 3.09 years, who was present for about half of the study time. Preschool B was situated in a semiresidential area of inner Belfast with a large immigrant population and was attended by a mixture of children who lived locally and those whose parents worked nearby. Hence, the children's cultural backgrounds were more diverse than in Preschool A: The majority came from (Northern) Irish families, but 4 of the children studied were Chinese, 1 was Malaysian, 1 Nigerian, and 1 Zimbabwean. In addition, 2 children were of mixed ethnicity (Irish/German and Irish/Spanish). The study group comprised 24 children (13 boys and 11 girls). Their age at the start of the study, which lasted for 2 months from April 2007, ranged from 3.23 to 4.65 years ($M = 4.11$).

Participant Observation

Participant observation allows a researcher to map out the social and cultural context of human activities, rather than taking an isolated reading of behavior as with more structured observational techniques (Hammersley & Atkinson, 2007; see Corsaro, 2003, on participant observation in preschool settings). This method was used to contextualize the quantitative research, partly by investigating the meaning and motivation of children's acts of tattling, and partly by getting a sense of how it felt, as an adult in a position of authority, to receive behavioral reports from children about their peers. The first author spent 2 months engaged in participant observation at Preschool A, working 3–4 days per week, for a total of 98 hr over 33 study days. He later spent 2 months at Preschool B, working 2–3 days per week, for a total of 96 hr over 23 study days. Both figures include time spent in quantitative sampling: After the first week, which was devoted solely to participant observation, most observation sessions started and finished with a period of participant observation. Field notes were typed up each afternoon after school had finished. In between the visits to the two preschools, and during the analysis phase after all observation was over, we analyzed the qualitative field notes and categorized emergent themes in the research (e.g., the various motivations for tattling).

Event Sampling

The quantitative method used most extensively in this study was the event sampling of children's reports of peers' behavior. Event sampling was chosen because it is a useful method for systematically recording a wide range of properties of everyday experience, and for generating novel hypotheses (Reis & Gable, 2000). At Preschool A, the first author spent 31 hr on event sampling, spread over 15 study days in continuous sessions of between 1 and 3 hr in duration. At Preschool B, he spent 35 hr on event sampling, over 15 study days. An event in this study was defined as a verbal report of a peer's behavior (whether counternormative or not), made by a child to someone else. Every event that was overheard by the observer was recorded on paper, as soon as possible after it occurred (usually within 1 min, and nearly always within 5 min). A small number of events in the sessions (< 20 in total) were omitted, either because of partial inaudibility or because the observer was too busy to record them when they occurred.

For each event, the following information was coded on preformatted coding sheets in the observer's notebook: the *time* at which the event occurred; the child(ren) who made the report (the *tattler*); the child(ren) who performed the reported action (the *miscreant*); the adult(s) or (occasionally) child to whom the report was made (the *audience*); the person most affected by the reported behavior (the *victim*), which was defined as either the tattler themselves, the audience, a third party, or nobody in particular; the *truth value* of the tattler's account (true, false, or indeterminate); the free-text *content* of the tattler's report (often verbatim); and a free-text description of the audience's *response* to the event. After data collection, content analysis of the content and response free-text fields was used to code data using the coding schemes in the two paragraphs below. For these additional categories, half of the events were rated independently by a second person who was blind to the observational predictions and was given the definitions below, along with examples similar to those in the appendices. Interrater reliability for content type was 90%, and for response type it was 89% (Cohen's kappas were .87 and .84, respectively).

Content type. Reports of peers' behavior were assigned to one of 10 categories, following Ross and den Bak-Lammers (1998). See Appendix A for examples. Eight of these categories were considered to represent various kinds of norm violation. Reports of *physical aggression* referred to any kind of unwanted physical contact, for example, hitting or pushing. *Property damage* reports described any kind of damage to property, for example, breaking a toy, or knocking over some blocks. *Property entitlement* covered reports of someone taking something that belonged to another child, or that another child had been using, as well as children refusing to share objects that they themselves owned or were using. A report of *social convention* referred to a violation of some conventional rule of the classroom, such as standing on a chair, or poor table manners. *Joint play violations* included reports of a child obstructing another child's play, or refusing to play alongside them. Reports of *taunting* referred to one child shouting at another or calling them names. *Deception* reports described another child lying to or otherwise misleading someone. Reports of *disagreement* referred to one child denying something that another child had said or believed in. There were also two categories of reports that did not meet the definition of tattling, as they did not describe counternormative behavior. In *nonjudgmental* reports the behavior described was innocent and the child who reported it did not seem to be seeking any punishment. Finally, *positive* reports consisted of an approving description of another child's prosocial behavior.

Response type. The actions of the audience in response to a behavioral report were assigned to one of eight categories, again following Ross and den Bak-Lammers (1998). See Appendix B for examples. *Supporting* a tattler involved intervening on his or her behalf, such as by verbally admonishing the miscreant, or compelling her to hand over a toy to another child. *Acknowledging* a report consisted of agreeing or sympathizing with the tattler but not saying anything to the miscreant about the reported behavior. *Excusing* an action involved asserting to the tattler that the behavior reported is innocent or justified. *Ignoring* was coded if the audience did not seem to respond to the tattler in any meaningful way. When *reprimanding* a tattler, the audience reproached the tattler for tattling or for the tattler's own reported behavior. In some cases, *both* tattler and miscreant were reprimanded—the implication being that they had both been involved in the reported activity—and this received its own code. Finally, a *questioning* response took place when the audience tried to find out exactly what had happened by questioning the tattler, the miscreant, or both, but was unable to reach any firm conclusions (i.e., the response did not resolve into any other category).

Point Sampling of Social Networks

Tattling is not just an individual activity but a relational activity between children and the peers whose behavior they report. It might therefore vary according to the social relations between individuals. We built up a picture of the social networks in the two classrooms and the sociability of individual children, by using point sampling to record the composition of children's play groups during free play sessions. In Preschool A, group membership was surveyed at a set time every morning. Several study days were also devoted to point sampling, taking a sample every 15 min for the duration of the morning. In all, 106 point samples were taken, each of which recorded the group membership of every child present. In the busier environment of Preschool B, point sampling was conducted on a more infrequent and ad hoc basis (18 samples, spread over nine study days).

Focal Follows and Dominance Analysis

In Preschool A, each child was observed individually for 1 hr. This hour was made up of one 30-min and two 15-min sessions per child, carried out on three separate days. Everything that the focal child did during these sessions was noted. The motivation for the focal follows was twofold: first, to check that the event sampling of the children's verbal reports was not biased toward the most audible or attention-grabbing children, and second, to investigate the dominance hierarchy within the classroom. A common method for analyzing the dominance hierarchies of toddlers is to code for incidences of direct aggression (e.g., Strayer & Strayer, 1976). However, examples of direct aggression become much less frequent by age 4 (Hawley, 1999) and were rarely observed in Preschool A; therefore, instead we analyzed several types of social interaction that included a clear element of direction of one child's behavior by another (cf. Barner-Barry, 1988; La Freniere & Charlesworth, 1983).

Child X was considered to have taken part in a dominant interaction with Child Y if any one of the following behavioral patterns occurred: (a) X initiated physical contact with Y, and Y did not resist; (b) X told Y to do something, and Y complied; (c) Y imitated X's behavior; (d) Y followed X to another part of the room. In each case, the behavior of both children was important, emphasizing the point that dominance is a function of dyadic relations rather than of individual behavior alone (Strayer & Strayer, 1976). However, for many dyads there were no examples of such interactions, and so it was impossible to construct a single transitive hierarchy that included all the children. Therefore, the ratio of total dominant to submissive interactions for each child was used as an index of relative dominance. The dominance hierarchy constructed using this measure was in accordance with the transitive subhierarchies that were constructed based on interactions between specific dyads, and with qualitative observations and informal teacher reports of children's relative dominance.

Teacher Ratings of Relational Aggression

Due to time constraints and the larger class size, and because we had some confidence from the first preschool that the event sampling procedure would not underestimate the level of tattling, no focal follows were carried out in Preschool B. Instead, we administered the Relational Aggression factor of the Preschool Social Behavior Scale–Teacher form (PSBS-T; see Crick, Casas, & Mosher, 1997) to the teacher and classroom assistant in Preschool B, to find out if there was a link between tattling and relational aggression. As mentioned in our Introduction, some forms of evaluative discourse in children have been analyzed as instances of relational aggression, defined as ''harming others through purposeful manipulation and damage of their peer relationships'' (Crick & Grotpeter, 1995, p. 711). The relational aggression factor of the PSBS–T consists of six descriptions of children's characteristic behavior (e.g., ''Tells others not to play with or be a peer's friend''), on which each child was rated using a 5-point Likert scale. Crick et al. (1997) demonstrated that this factor was internally consistent and independent of the overt aggression, prosocial behavior, and depressed affect factors. McEvoy, Estrem, Rodriguez, and Olson (2003) found strong intermethod agreement between teacher ratings on the PSBS–T, peer nominations, and direct observation of relational aggression.

Results

Content of Children's Behavioral Reports

Children's tattling, operationalized as the reporting of the various categories of norm violations described in the Method section, was far more frequent than the reporting of positive or neutral behavior. Event sampling recorded 354 examples (93.1%) of tattling ($M = 1.26$ reports per child per day attended, $SD = 1.12$), 25 examples (6.6%) of nonjudgmental talk ($M = 0.09$, $SD = 0.14$), and just 1 example (.3%) of positive talk about a peer's activities. There were no significant differences between the preschools either in the mean frequency of all behavioral reports per child, or in the mean proportion of negative reports, both $ts < 1$, both $ps > .5$. The range of behavioral reporting was 0–6.31 events per child per day attended.

Tattling might be expected to be more noticeable, and thus more readily observable, than other forms of talk about peers. However, a similar effect was observed in the data from the focal follows, where it would have been difficult to miss any kind of behavioral report about a third party made by the focal child. Of 32 instances in the focal follow data where a focal child was involved in a behavioral report, only two reports (6.3%) were nonjudgmental, and none were positive. Examples of actual tattling content are listed in Appendix A. No examples were found of one of the predefined categories of tattling content (deception). For the other seven categories, mean frequencies of children's tattling are shown in Table 1. A series of t tests revealed no significant differences in tattling content between preschools, though there were trends for children in Preschool A more frequently to report both social conventional issues, $t(16.2) = 1.86$, $p = .081$, and disputes over property entitlement, $t(38) = 1.74$, $p = .090$.

Truthfulness

For event samples where it was possible to determine unambiguously the truth of a behavioral

Table 1
Frequency of Reported Categories of Counter-normative Behavior

	Preschool A M^a (*SD*)	Preschool B M^a (*SD*)	Overall M^a (*SD*)
Property entitlement	0.53 (0.49)	0.32 (0.29)	0.41 (0.39)
Physical aggression	0.25 (0.27)	0.38 (0.40)	0.33 (0.36)
Social convention	0.31 (0.47)	0.09 (0.11)	0.18 (0.32)
Joint play	0.08 (0.21)	0.17 (0.20)	0.14 (0.21)
Property damage	0.16 (0.23)	0.06 (0.16)	0.10 (0.20)
Taunting	0.07 (0.10)	0.12 (0.15)	0.10 (0.14)
Disagreement	0.02 (0.05)	0.01 (0.03)	0.01 (0.04)

[a] Mean frequency of events reported by each child, per standardized school day attended.

report, a mean of 90.0% of reports were found to be true. As many children in each preschool were never observed to make a false report, there was no difference whatsoever between the two schools on this measure, $t(32) = 0.020$, $p = .99$. However in 43.1% of event samples, on average, the truth value of the report was undetermined. This was partly because the busy nature of the classroom environment sometimes made it difficult to record all the relevant antecedents of the children's disputes, and partly because some reports (e.g., "He's not listening to me!") were difficult to assign objective truth values to without intensive questioning of the children involved. In the focal follow data, where it was much easier to determine the truth of children's claims, only 9 of 32 (28.1%) reports were of undetermined truth value, and no false reports were recorded. In most of the undetermined cases, it is unlikely that the reports were false, as the alleged miscreants rarely denied their offenses, and the teachers rarely accused tattlers of lying. Moreover, the proportions of *deceptive* reports in our study were well below the 10% of *false* reports that were recorded. For example, Child *B* might crash into Child *A* on a tricycle, and Child *A* might falsely report to the teacher that Child *C* had crashed into him. This might be an error in face recognition, or a slip of the tongue: There were very few events in which the observer suspected that deliberate fabrication of a report was occurring. Mistaken ascriptions of intention (e.g., "She pushed me!" when one child accidentally knocked into another) were more common.

Audience Responses to Tattling

Children were far more likely to report the behavior of their peers to adults than to other children: In the event samples, only 4 of 380 events (1.1%) were reports made to other children. Table 2 shows the mean shares of the various types of audience response to tattling, ordered by overall frequency. Examples of actual responses are supplied in Appendix B. The most common response was supporting the tattler, which accounted for around 50% of responses. If acknowledging is included as a favorable response for the tattler, as it constitutes positive attention from an authority figure, almost 70% of responses were favorable for the tattler, and only 6% were clearly unfavorable (a reprimand for the tattler alone or for both tattler and miscreant). There were no significant differences between the preschools in the proportions of these responses, all ts < 1.5, all ps > .1.

Table 2
Proportions of Audience Responses to Tattling

	Preschool A M^a (*SD*)	Preschool B M^a (*SD*)	Overall M^a (*SD*)
Supporting	0.52 (0.27)	0.44 (0.19)	0.47 (0.23)
Acknowledging	0.17 (0.27)	0.25 (0.24)	0.22 (0.25)
Excusing	0.08 (0.09)	0.12 (0.13)	0.11 (0.11)
Ignoring	0.14 (0.17)	0.07 (0.09)	0.10 (0.13)
Reprimanding	0.03 (0.05)	0.03 (0.09)	0.03 (0.08)
Both reprimanded	0.04 (0.13)	0.02 (0.04)	0.03 (0.09)
Questioning	0.12 (0.26)	0.13 (0.13)	0.12 (0.19)

[a] Mean share of responses for each child who was tattling.

Egocentric and Sociocentric Reports

Tattling was highly egocentric, in that it tended to focus on achieving help or punishment for another child's negative behavior toward the tattler (mean share = 0.77, $SD = 0.24$). Tattling on behalf of third parties was quite rare ($M = 0.06$, $SD = 0.11$), and reporting a transgression that directly affected the audience was more unusual still ($M = 0.01$, $SD = 0.02$). Tattling where there was no clear victim—as with most breaches of social convention—was more common ($M = 0.16$, $SD = 0.17$). Children in Preschool A were more likely than in Preschool B to report incidents where there was no clear victim, $t(34) = 2.17$, $p = .037$, which might be linked to their tendency to report more social conventional violations.

Dominance

In Preschool A there was a strong positive correlation between children's dominance, as measured by the ratio of their dominant to subordinate interactions, and the rate at which they tattled on other children, $r = .881$, $N = 16$, $p = .000007$. That is, dominant children reported others' transgressions significantly more than subordinate children. One possible explanation for this is that children who talked to the teacher more often were more likely to engage in tattling. However, there were no correlations between overall rates of addressing an adult in the classroom, as recorded in the focal follows, and either rates of tattling or ratio of dominant to subordinate interactions, both rs < .25, both ps > .3. Alternatively, sociability might be expected to increase tattling on the part of dominant children, as more social interactions might lead to more potential for conflicts and therefore more opportunities to report those conflicts. As the average number of play partners for a given child would be affected by total attendance on the days that he or she was present, we calculated an index of sociability for each child, based on the ratio between the actual number of interactions with every other child and the potential number of interactions that could have taken place (i.e., the number of point samples at which both children were present). There was no correlation between this sociability index and either the frequency of children's tattling or the ratio of dominant to subordinate interactions, both $r < .3$, both $p > .3$. Hence, there seemed to be a specific link between dominance and tattling in this preschool. (As explained in the Method section, dominance was not analyzed in Preschool B.)

Relational Aggression

In Preschool B there was a strong correlation between tattling frequency and a child's score on the relational aggression section of the PSBS–T teacher-rated questionnaire, $r = .590$, $N = 24$, $p = .002$. This demonstrates that tattling in this preschool covaried with verbal indices of relational aggression. (The PSBS–T questionnaire was not administered in Preschool A, as explained in the Method section.)

Gender Effects

No specific gender effects were predicted. Boys ($M = 3.21$, $SD = 0.88$, $n = 13$) were far more likely than girls ($M = 1.60$, $SD = 0.72$, $n = 11$), to be seen by the teacher as relationally aggressive, $t(22) = 4.84$, $p = .00008$. Girls' behavioral reports were slightly more truthful than boys' reports, but both were highly truthful ($M = 0.98$, $SD = 0.05$, $n = 19$ for girls; $M = 0.84$, $SD = 0.26$, $n = 21$ for boys), and this effect was not particularly significant, $t(19.8) = 2.36$, $p = .029$, given the number of potential gender effects that were tested (Bonferroni-corrected $\alpha = 0.0025$). As there was no difference in the overall tattling rates between genders, $t(38) = .882$, $p = .383$, it was possible to directly compare the frequencies of tattling by boys and girls on the various categories of norm violation (see Figure 1). The only significant difference was in tattling on physical aggression, which boys ($M = 0.49$ per day, $SD = 0.41$, $n = 21$) reported much more frequently than girls ($M = 0.15$ per day, $SD = 0.16$, $n = 19$), $t(26.7) = 3.46$, $p = .002$. There was no significant effect of tattler gender on the frequency of any of the various types of response to tattling, all $ts < 1.5$, all $ps \geq .15$. No effects of tattler gender were found on the frequency of behavioral reports, proportion of negative reports, or the likelihood of the tattler being the victim, all $ts < 1.0$, all $ps > .3$.

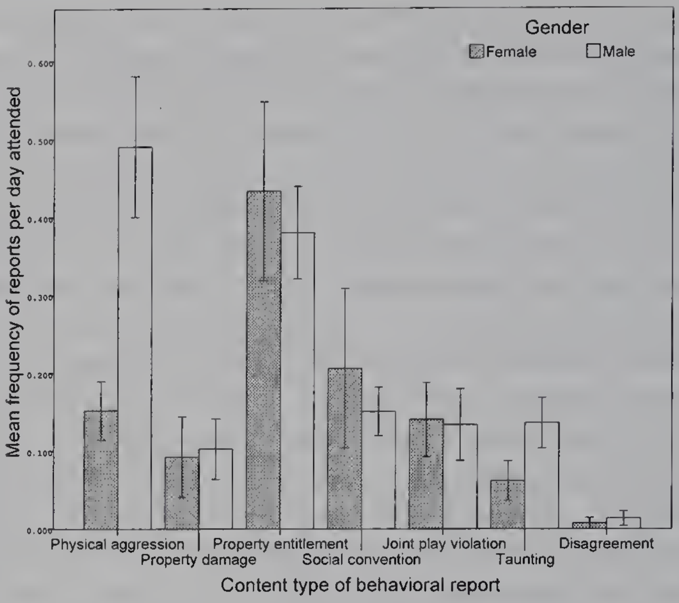


Figure 1. Mean frequencies of tattling (adjusted for attendance) by boys and girls in both preschools on the various categories of norm violation. Error bars represent ±1 *SE*.

Discussion

Our data suggest that tattling is an important form of social communication for many young children. The great majority of children's talk about their peers' behavior took the form of descriptions of norm violations. This was a robust finding that held across both sexes in both preschools, and also in the focal follow data (for which sampling bias was less of an issue). Moreover, our data reiterate the finding of a bias toward talk about counternormative behavior among all four groups of siblings studied by Ross and den Bak-Lammers (1998). Taken together, our preschool study and their sibling study demonstrate that a bias toward reporting norm violations by other children—while not necessarily universal—is present in children of various ages in differing social contexts. It seems that when children first start talking to adults about what another child has done, they use this faculty largely to report behavior of which they disapprove. More generally, Miller and Sperry (1988) found that 2-year-olds' stories about past events were highly evaluative, and that 64% of stories were concerned with negative experiences. This may reflect an early bias toward sharing negative experiences with caregivers, perhaps caused by an "ontogenetic adaptation" (Bjorklund & Pellegrini, 2000) for children to seek aid or sympathy from adults when suffering a negative affective response. However, the extent to which such a bias is recruited in everyday life and the precise kinds of events that are reported are likely to be affected by local cultural norms of discourse (Burger & Miller, 1999; Miller, Cho, & Bracey, 2005).

Children in the present study were much more likely to report disputes arising from issues of property entitlement or physical aggression than they were to report joint play disputes, taunting, property damage, simple disagreements, or deception. Again, this is consistent with the findings of Ross and den Bak-Lammers (1998). Nucci and Turiel (1978) likewise found that preschool children were more likely to protest about moral (harm-related) than conventional violations in a classroom setting. Events involving physical harm are also common in 2-year-olds' stories about the past (Miller & Sperry, 1988). It seems that young children may preferentially report breaches of "affect-backed" norms (see Nichols, 2004), involving physical harm or property loss. Even at the age of 3, nonetheless, they can generalize this form of communication to breaches of simple social norms (such as shouting, or going into inappropriate areas of the classroom) but very rarely report instances of deception or disagreement—perhaps because an understanding of such complex linguistic norms requires an advanced theory of mind (Astington, 2006).

Regardless of content, the great majority (90%) of children's reports about peers' behavior were true, at least for cases in which truth could be easily determined. This figure is similar to Ross and den Bak-Lammers's (1998) finding of a mean of 94.3% truthful reports across all four groups that they studied. Moreover, most false reports seemed to be mistaken rather than fabricated. This bias toward truthful communication is interesting because children of this age—and younger—are certainly capable of lying in both naturalistic and experimental settings (for a review, see Lee & Talwar, in press). Indeed, children in the present study were often observed to lie in response to tattling, usually by means of a simple denial such as "No I didn't!" As strategic deception requires a high level of executive control (Hala & Russell, 2001), it may be that by this stage of development children's executive competence enables them to make false denials of responsibility, but not yet to fabricate complex narratives that might successfully absolve them of responsibility. This finding argues against stereotypical views of tattling as "telling tales" (i.e., making up stories), suggesting as it does that young children's reports of peers' behavior are usually reliable.

Children were more inclined to report the behavior of those who had harmed or offended them directly. At a pragmatic level, this finding may reflect a general egocentric bias in children's discourse (Rubin, 1973). At a motivational level, it may also reflect a greater incentive to punish peers who have offended against the subject rather than those who have harmed a third party. Nevertheless, children as young as 2 are capable of sociocentric speech (Bruner, 1983), and indeed in about 25% of cases, in the present study, children reported social conventional violations for which there was no clear victim. It is interesting that children reported social conventional violations much more frequently than transgressions against a third party. This suggests a certain level of social cognitive awareness, in that they may have assumed that the victims in the latter type of case could speak up for themselves.

Tattling was rarely ignored by adults and often led to a teacher or classroom assistant intervening in support of a child—whether by punishing the reported miscreant or by resolving a dispute in favor of the tattler. There would seem to be little risk involved in tattling, in that few tattling events led to a negative response for the speaker (this is also true of the family context; Ross & den Bak-Lammers, 1998). Furthermore, qualitative observations during the participant observational part of the current study showed that while an explicit threat of tattling sometimes deterred peers from carrying out an undesired activity, tattling rarely led to aggressive retaliation on the part of peers (except in the form of reciprocal tattling). Compared to direct physical retaliation—which might lead either to an escalation in aggression or to punishment by staff—tattling might thus be a useful, and low-risk, social strategy for young children to follow in seeking punishment for a perceived wrong. Yet qualitative observations also indicated that children's motivations for tattling were highly variable across individuals and situations (see Ingram, 2009, for a detailed analysis). Punishment did not result in the majority of cases and was usually mild: Admonishment sometimes consisted simply of saying the transgressor's name in a disapproving tone. Often it seemed to be enough for children that an adult was paying attention to them and that their concerns about another's behavior were being acknowledged. Tattling is therefore more likely to be motivated by a general drive to externalize emotional problems, than by a conditionally learned strategy to achieve punishment.

There were notable individual differences in tattling. Analysis of social hierarchies in Preschool A showed that dominant individuals tattled significantly more frequently than other children. Tattling is sometimes a response to a (perceived) aggressive action by another individual, and this may mean that it is practiced more by dominant children, who more often retaliate against peers' aggression (Strayer & Strayer, 1976). In an environment where children's physical aggression is frequently punished by adults, dominant individuals' retaliatory impulses may be socialized and directed into indirect, verbal behavior (Cooney et al., 1996; Hawley, 1999). Tattling may be one of several interpersonal strategies—including relationally aggressive behavior such as saying "I'm not your friend," verbally aggressive behavior such as threats or taunts, and direct physical aggression such as pushing—which some preschool children use, to varying degrees, to achieve social dominance. Although dominant children were no more likely than other children to address the teachers in general terms, it could also be that dominant children were less shy about the specific activity of drawing an adult's attention to social problems. These possibilities are not mutually exclusive; that is, manipulating others' attention may contribute to dominance in young children.

The results of the PSBS–T questionnaire administered in Preschool B suggest that there are links between tattling and relational aggression. The same children who often engaged in relationally aggressive activities, such as telling peers that they would not be their friend, tended to tattle the most frequently. On the other hand, relational aggression has been defined as behavior that "harms others through damage to their peer relationships (e.g., using social exclusion or rumor spreading as a form of retaliation)" (Crick et al., 1997, p. 579). Relational aggression is often also characterized as covert (Archer & Coyne, 2005)—for example, whispering behind someone's back—whereas tattling in these two preschools was overt and unashamed, often taking place right in front of the tattling target. As it is often overt, tattling may be more closely related to social aggression, a construct defined to include both covert and overt forms of nonphysical aggression (Archer & Coyne, 2005; Galen & Underwood, 1997). Moreover, it is unlikely that children who tattled always intended to *harm* the tattling targets. Sometimes, they may well have been motivated by a desire to win praise from the teacher or to correct some inequity in the classroom (Ingram, 2009), in an early form of restorative justice. More research is needed on the affective dimensions of tattling, and on its relation with other forms of social communication.

One advantage of doing participant observation, in addition to quantitative sampling, was that it gave us a sense of the broad range of motivations for tattling. Tattling could sometimes be a strategic element of interchild conflict, as demonstrated by the fact that children occasionally threatened to do it ("I'm telling the teacher!") as a sanction against unwelcome behavior by a peer. Yet tattling was not always an aggressive reprisal for a perceived wrong against the tattler: It sometimes seemed to be more concerned with providing useful information about rule breaking to an authority figure. When children tattled about nonegocentric events, it often seemed to be the principle of justice that most concerned them—an assertion of the rights and wrongs of the matter, rather than any particular settlement. For example, a frequent topic of tattling occurred when one child was perceived to take more than his or her fair share of bread or milk at the snack table. Other children would report this even when they themselves had plenty of food and drink in front of them. They seemed to be concerned with the principle of one child having relatively more than the others, rather than with any absolute loss of resources on their own part (for related experimental results, see Damon, 1977; Fehr, Bernhard, & Rockenbach, 2008). Hence the motivating force behind tattling may sometimes be a drive to enforce the rules as the child perceives them, in pursuit of a sense of justice or social equity. What might produce such a drive? One cause might be a reluctance to see another go unpunished for an offense for which the tattler has been punished in the past, thus reducing the tattler's relative standing. Another cause, which would fit well with our dominance results, might be a tendency for leaders to gain prestige from enforcing social norms (O'Gorman, Henrich, & Van Vugt, 2009).

There were several limitations to the present study. It was based on hand-written observations of a large body of children's discourse: An analysis based on several hours of audio or video recording of focal children would contain fewer examples of tattling but might allow for finer-grained observation of the motivational and affective dimensions of tattling. Although two preschools were compared—and tattling took a generally similar form in each, with no robustly significant differences—both schools were in the same city, and so it would be interesting to contrast the results from these schools with results from culturally different populations. It would also be worth comparing preschoolers' reporting of peer behavior with similar activities by older children in educational settings, thus complementing and extending the longitudinal study of Ross and den Bak-Lammers (1998) in home settings. It would be especially useful to carry the study of tattling up to preadolescence, as this may be when the activity first becomes derogated by peers (cf. Friman et al., 2004).

In this article we have extended the study of tattling to preschool settings, showing that it took a similar form in this context as for preschool-age children in the home context. Our results in the areas of negative bias, truthfulness, and adult responses to tattling are broadly in agreement with the findings of Ross and den Bak-Lammers (1998) for tattling between siblings. Children's general bias toward reporting norm violations also recalls the sensitivity toward norm violations in experimental settings reported by Harris and Núñez (1996) and Rakoczy et al. (2008). Further experimental work needs to be done regarding the cognitive mechanisms responsible for this bias: Are norm violations simply more salient, and therefore more worthy of verbal attention, or are children's verbal accounts of behavior always influenced by strategic considerations? Individual variations in tattling, and in particular the relations that we found with dominance and relational aggression, require carefully designed follow-up research. We conclude that tattling is a robust but individually flexible pattern of behavior, which is likely to undergo interesting developmental changes as children grow older and acquire an advanced theory of mind (see Ingram, Piazza, & Bering, 2009). Tattling may be an active part of the process by which some children learn to live in a world of social norms and social conflicts. It is thus an early example of the human ability to use language to externalize strategic problems and draw them into the public sphere.

References

Archer, J., & Coyne, S. M. (2005). An integrated review of indirect, relational, and social aggression. *Personality and Social Psychology Review*, *9*, 212–230.

Astington, J. W. (2006). The developmental interdependence of theory of mind and language. In N. J. Enfield & S. C. Levinson (Eds.), *Roots of human sociality* (pp. 179–206). Oxford, UK: Berg.

Barner-Barry, C. (1988). The structure of politically relevant behaviours in pre-school groups. In M. R. A. Chance (Ed.), *Social fabrics of the mind* (pp. 261–276). Hove, UK: Erlbaum.

Bjorklund, D. F., & Pellegrini, A. D. (2000). Child development and evolutionary psychology. *Child Development*, *71*, 1687–1708.

Brewer, G. A., & Selden, S. C. (1998). Whistle blowers in the federal civil service: New evidence of the public service ethic. *Journal of Public Administration Research and Theory*, *8*, 413–439.

Bruner, J. S. (1983). *Child's talk: Learning to use language*, Oxford, UK: Oxford University Press.

Burger, L. K., & Miller, P. J. (1999). Early talk about the past revisited: Affect in working-class and middle-class children's co-narrations. *Journal of Child Language*, *26*, 133–162.

Cameron, D. (1997). Performing gender identity: Young men's talk and the construction of heterosexual masculinity. In S. Johnson & U. H. Meinhof (Eds.), *Language and masculinity* (pp. 47–64). Oxford, UK: Blackwell.

Cooney, M., Hutchison, L., & Costigan, V. (1996). From hitting and tattling to communication and negotiation: The young child's stages of socialization. *Early Childhood Education Journal*, *24*, 23–27.

Corsaro, W. A. (2003). *"We're friends, right?" Inside kids' cultures*. Washington, DC: Joseph Henry Press.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task *Cognition*, *31*, 187–276.

Crick, N. R., Casas, J. F., & Mosher, M. (1997). Relational and overt aggression in preschool. *Developmental Psychology*, *33*, 579–588.

Crick, N. R., & Grotpeter, J. K. (1995). Relational aggression, gender, and social-psychological adjustment. *Child Development*, *66*, 710–722.

Damon, W. (1977). *The social world of the child*. San Francisco: Jossey-Bass.

Den Bak, I. M., & Ross, H. S. (1996). "I'm telling!" The content, context, and consequences of children's tattling on their siblings. *Social Development*, *5*, 292–309.

Dunn, J. (1988). *The beginnings of social understanding*. Oxford, UK: Basil Blackwell.

Dunn, J. (2006). Moral development in early childhood and social interaction in the family. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 331–350). Mahwah, NJ: Erlbaum.

Dunn, J., & Munn, P. (1987). Development of justification in disputes with mother and sibling. *Developmental Psychology*, *23*, 791–798.

Engel, S., & Li, A. (2004). Narratives, gossip, and shared experience: How and what young children know about the lives of others. In J. M. Lucariello, J. A. Hudson, R. Fivush, & P. J. Bauer (Eds.), *The development of the mediated mind*. (pp. 151–174). Mahwah, NJ: Erlbaum.

Fehr, E., Bernhard, H., & Rockenbach, B. (2008). Egalitarianism in young children. *Nature*, *454*, 1079–1083.

Fine, G. A. (1977). Social components of children's gossip. *Journal of Communication*, *27*, 181–185.

Fine, G. A. (1986). The social organization of adolescent gossip: The rhetoric of moral evaluation. In J. Cook-Gumperz, W. A. Corsaro, & J. Streeck (Eds.), *Children's worlds and children's language* (pp. 405–423). Berlin, Germany: Mouton de Gruyter.

Friman, P. C., Woods, D. W., Freeman, K. A., Gilman, R., Short, M., McGrath, A. M., et al. (2004). Relationships between tattling, likeability, and social classification: A preliminary investigation of adolescents in residential care. *Behavior Modification*, *28*, 331–348.

Galen, B. R., & Underwood, M. K. (1997). A developmental investigation of social aggression among children. *Developmental Psychology*, *33*, 589–600.

Goodwin, M. H. (1990). *He-said-she-said: Talk as social organization among black children*. Bloomington: Indiana University Press.

Hala, S., & Russell, J. (2001). Executive control within strategic deception: A window on early cognitive development? *Journal of Experimental Child Psychology*, *80*, 112–141.

Hammersley, M., & Atkinson, P. (2007). *Ethnography: Principles in practice* (3rd ed.). London: Routledge.

Harris, P. L. (2000). *The work of the imagination*. Oxford, UK: Blackwell.

Harris, P. L., & Núñez, M. (1996). Understanding of permission rules by preschool children. *Child Development*, *67*, 1572–1591.

Harris, P. L., Núñez, M., & Brett, C. (2001). Let's swap: Early understanding of social exchange by British and Nepali children. *Memory and Cognition*, *29*, 757–764.

Hawley, P. H. (1999). The ontogenesis of social dominance: A strategy-based evolutionary perspective. *Developmental Review*, *19*, 97–132.

Hurlock, E. B., & McDonald, L. C. (1934). Undesirable behavior traits in junior high school students. *Child Development*, *5*, 278–290.

Ingram, G. P. D. (2009). *Young children's reporting of peers' behaviour*. Unpublished doctoral dissertation, Queen's University Belfast.

Ingram, G. P. D., Piazza, J. R., & Bering, J. M. (2009). The adaptive problem of absent third-party punishment. In H. Høgh-Olesen, P. Bertelsen, & J. Tønnesvang (Eds.), *Human characteristics: Evolutionary perspectives on human mind and kind* (pp. 205–229). Newcastle-upon-Tyne, UK: Cambridge Scholars.

La Freniere, P., & Charlesworth, W. R. (1983). Dominance, attention, and affiliation in a preschool group: A nine-month longitudinal study. *Ethology and Sociobiology*, *4*, 55–67.

Lee, K., & Talwar, V. (in press). *Children and lying: A century of scientific research*. Chichester, UK: Wiley-Blackwell.

Loftus, E. F. (1979). *Eyewitness testimony*. Cambridge, MA: Harvard University Press.

McConnell, G. (1963). Behavior that annoys teachers. *The Elementary School Journal*, *63*, 448–452.

McEvoy, M. A., Estrem, T. L., Rodriguez, M. C., & Olson, M. L. (2003). Assessing relational and physical aggression among preschool children: Intermethod agreement. *Topics for Early Childhood Special Education*, *23*, 51–61.

Mettetal, G. (1983). Fantasy, gossip, and self-disclosure: Children's conversations with friends. *Communication Yearbook*, *7*, 717–735.

Miller, P. J., Cho, G. E., & Bracey, J. R. (2005). Working-class children's experience through the prism of personal storytelling. *Human Development*, *48*, 115–135.

Miller, P. J., & Sperry, L. L. (1988). Early talk about the past: The origins of conversational stories of personal experience. *Journal of Child Language*, *15*, 293–315.

Morrison, J. Q., & Jones, K. M. (2007). The effects of positive peer reporting as a class-wide positive behavior support. *Journal of Behavioral Education*, *16*, 111–124.

Nichols, S. (2004). *Sentimental rules: On the natural foundations of moral judgment*. Oxford, UK: Oxford University Press.

Nucci, L. P., & Turiel, E. (1978). Social interactions in the context of moral and conventional transgressions. *Child Development*, *49*, 400–407.

O'Gorman, R., Henrich, J., & Van Vugt, M. (2009). Constraining free riding in public goods games: Designated solitary punishers can sustain human cooperation. *Proceedings of the Royal Society B*, *276*, 323–329.

Piaget, J. (1932). *The moral judgment of the child* (M. Gabain Trans.). London: Kegan Paul, Trench, Trubner & Co.

Rakoczy, H., Warneken, F., & Tomasello, M. (2008). The sources of normativity: Young children's awareness of the normative structure of games. *Developmental Psychology*, *44*, 875–881.

Reis, H. T., & Gable, S. L. (2000). Event-sampling and other methods for studying everyday experience. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 190–222). Cambridge, UK: Cambridge University Press.

Rosenfeld, R., Jacobs, B. A., & Wright, R. (2003). Snitching and the code of the street. *British Journal of Criminology*, *43*, 291–309.

Ross, H. S., & den Bak-Lammers, I. M. (1998). Consistency and change in children's tattling on their siblings: Children's perspectives on the moral rules and procedures of family life. *Social Development*, *7*, 275–300.

Rubin, K. H. (1973). Egocentrism in childhood: A unitary construct? *Child Development*, *44*, 102–110.

Skinner, C. H., Cashwell, T. H., & Skinner, A. L. (2000). Increasing tootling: The effects of a peer-monitored group contingency program on students' reports of peers' prosocial behaviors. *Psychology in the Schools*, *37*, 263–270.

Smith, M. E. (1932). The preschool child's use of criticism. *Child Development*, *3*, 137–141.

Strayer, F. F., & Strayer, J. (1976). An ethological analysis of social agonism and dominance relations among preschool children. *Child Development*, *47*, 980–989.

Sugiyama, L. S., Tooby, J., & Cosmides, L. (2002). Cross-cultural evidence of cognitive adaptations for social exchange among the Shiwiar of Ecuadorian Amazonia. *Proceedings of the National Academy of Sciences of the United States of America, 99*, 11537–11542.

Tashakkori, A., & Teddlie, C. (Eds.). (2003). *Handbook of mixed methods in social and behavioral research*. Thousand Oaks, CA: Sage.

Thompson, R. A. (1994). Emotion regulation: A theme in search of definition. *Monographs of the Society for Research in Child Development, 59*, 25–52.

Turiel, E. (2006). Thought, emotions, and social interactional processes in moral development. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 7–35). Mahwah, NJ: Erlbaum.

Williams, M. (1989). You can curtail tattling. *Learning, 17*, 53.

Appendix A

Example Content Types of Children's Behavioral Reports

Report category	Examples of report content
Physical aggression	"He tried to bited [sic] me" "Jack hurt me with the wheel"
Property damage	"She hit this thing" "Matthew, he's knocking down the tower"
Property entitlement	"He won't let me have the scissors" "Mudiwa took two breads"
Social convention	"Both of them 'uns have Power Rangers—I seen them under the table" "They're being very loud"
Joint play violation	"She's not finished yet [with the paintbrush]; she won't hurry up" "I wanted to be in there [the box] and Jack pushed it off me"
Taunting	"She's shouting at me" "James said I'm not Superman; James said I'm bad"
Disagreement	"She said she was in a different school" "I stood on Adam's foot—he said I did it on purpose but I didn't"
Nonjudgmental	"He didn't get any milk" "He was trying to get in the car and I jumped in first"
Positive	"Emma gave me a flag"

Appendix B

Examples of Responses to Tattling by Audience

Response category	Examples of response content
Supporting	Confiscated toys: "You know you're not allowed them" "Tell him I'm not happy about that"
Acknowledging	Said "Does she?" and looked at the dolls Told tattler that she would talk to miscreant about it (but never seemed to)
Excusing	"That's OK" "It was probably an accident"
Ignoring	Pretty much ignored (probably because the boys had been throwing things too) More or less ignored, because everyone was going down into the playground
Reprimanding tattler	Told tattler not to be fighting "You are all little tell-tales, aren't you—always telling tales on each other!"
Both tattler and miscreant reprimanded	Ascertained that it was an accident, and told both boys to say sorry to each other Called both boys over and told them sternly not to climb on things
Questioning	Asked if it was an accident Questioning interrupted by arrival of tattler's mother

Child Development, May/June 2010, Volume 81, Number 3, Pages 958–971

Conversation, Development, and Gender: A Study of Changes in Children's Concepts of Punishment

Patrick J. Leman
Royal Holloway, University of London

Marina Björnberg
Linköping University

One hundred thirty-three children, average age 9.58 years, were read vignettes describing a moral transgression and then discussed what would constitute a fair punishment. Children's gender influenced conversation dynamics (use of interruption and other simultaneous speech) but not conversation content (justifications for judgments). All children who engaged in conversation showed development in judgments after 8 weeks compared with a control (no interaction) group. However, there was no relation between conversation dynamics or content and development, or any effects of gender, on the developmental process. The benefits of peer collaboration were general rather than specific to the stimulus story.

Despite abundant evidence that peer conversation promotes learning and development, much remains to be known about the features of interaction that are associated with changes in cognition. For instance, although the quality and type of arguments used in discussion have been linked with advances in reasoning in some studies (e.g., Felton & Kuhn, 2001; Psaltis & Duveen, 2007), other studies have found no links between argument content, quality, or frequency and learning (Fernandez-Berrocal & Santamaria, 2006; Howe, Tolmie, & Rodgers, 1992). It is important to build a clear understanding of the features of conversation that promote development because if we can identify what makes a conversation lead to development, we can begin to devise ways of making peer collaboration more productive.

Our central research questions concern the relations among gender, conversation, and development. These more general research questions fall, broadly, into three separate sets of hypotheses: first, predictions relating to the links between gender and conversation; second, predictions about the link between conversation (peer interaction) and development; and third, predictions about how features of conversation may link with development. By focusing on gender, we are exploring a key aspect of children's social identity that is known to influence conversational dynamics and has been proposed as a source of variation in moral judgment and reasoning. In terms of conversation, we explore how the content and dynamics of children's conversations may link with longer term change in judgments. The present study examines children's conversations and judgments about what constitutes a fair (or just) punishment.

The development of an idea of punishment was first studied by Piaget (1932), but aside from a few studies (e.g., Brinker, Goldstein, & Tisak, 2003; Saunders & Goddard, 2007; Tisak, 1993), the question of how children conceptualize different forms of punishment has subsequently received little attention from researchers. This neglect is surprising because what constitutes a fair punishment is something to which children pay a great deal of attention, and anecdotal evidence suggests that it is a frequent topic of discussion with parents and peers.

Piaget distinguished two forms of punishment: *expiatory punishment* and *punishment by reciprocity*. Expiatory punishments focus on retribution or the severity of punishment as the best means of addressing any wrongdoing. Younger children often judge expiatory punishment as the most fair. Older children, in contrast, often judge punishment by reciprocity as fairer. This form of

---

Part of this research was conducted to fulfill the requirements of the second author's master's thesis at Linköping University, Sweden. We are grateful to the schools, pupils, and staff who helped make the study a success. We also thank Sophie Purnell for her assistance in collecting the data, and two anonymous reviewers for their feedback on an earlier version of this manuscript. The research reported here was supported by a British Academy Senior Research Fellowship to the first author.

Correspondence concerning this article should be addressed to Patrick J. Leman, Department of Psychology, Royal Holloway, University of London, Egham, Surrey, United Kingdom. Electronic mail may be sent to patrick.leman@rhul.ac.uk.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8103-0021

punishment focuses on restorative aspects of punishment and putting right what was done wrong. According to Piaget, most children shift from a preference for expiatory to reciprocal forms of punishment between around 8 and 10 years. In the present study, we used the distinction between different forms of punishment as a trigger for 9-year-old children's conversations, and to gauge whether there were changes in children's punishment orientation (expiatory or reciprocal) after conversation.

Gender and Conversation

Gender has a profound influence on the dynamics of children's conversations. Examining the role of gender in conversation and development is important because if gender affects the dynamics of interaction then this may impact on the ideas and information that are introduced to a discussion and may also affect how far ideas are discussed and debated once they have been introduced. If, for instance, one child continually interrupts another in a discussion an important piece of information may be missed or an explanation may not be adequately scrutinized.

Boys' conversations are more dominance oriented, and involve more features such as interruption, whereas girls' conversations are more collaborative (see Leaper & Smith, 2004). Gender differences in conversations are evident from at least 7 years of age in both spontaneous interactions (Leaper, Tenenbaum, & Shaffer, 1999), classroom exchanges (Nelson-Le Gall & DeCooke, 1987), and more formal, problem-solving discussions (Leman, Ahmed, & Ozarow, 2005). In most contexts boys are more assertive and girls more affiliative in terms of conversation dynamics. Additionally, boys use more simultaneous speech acts than girls. Simultaneous speech acts include attempts to take the conversational "floor" such as interruptions and overlapping speech (Coates, 1989). Leman et al. (2005) found that the gender differences in the use of simultaneous speech acts were particularly pronounced in cross-gender (boy–girl) interaction.

Our first two hypotheses were derived from this previous work charting the influence of gender on children's conversations. Specifically, our first hypothesis predicted that boys will use more simultaneous speech acts (interruptions) than girls. Our second hypothesis predicted greater differences between genders in mixed than in same-sex interactions (i.e., boys will interrupt more than girls generally, but this difference will be especially marked in boy–girl pairs).

Gender and Moral Judgment

There has been considerable debate about whether there are gender differences in moral reasoning. Notably, Gilligan (1982) has argued that women and girls reason using an ethic of care with a focus on relationships, whereas men and boys tend to make more abstract, justice-based moral judgments. In conversation and discussion one might anticipate that these different moral orientations would lead to differences of opinion or even to conflict when children seek to justify their judgments. Even children who agree on the same moral judgment might disagree over the reasons for it.

Although the suggestion of gender differences in moral reasoning and orientation has received considerable research interest, review studies find no consistent gender differences in moral orientation in adults (e.g., Jaffee & Hyde, 2000; Walker, 1984, 2006). So in the present study, we did not expect gender to be a significant variable in terms of the choice of punishment (reciprocal or expiatory) or in the types of argument that children used to justify their choices.

Social Interaction and Development

Doise, Mugny, and Perret-Clermont (1975) demonstrated that children working in groups showed advances in conservation ability compared with those not involved in interaction. Subsequently, Doise and Mugny (1984) suggested that sociocognitive conflict—that is, the social manifestation of differences in perspective and understanding between children—is a vital driver for development. Ames and Murray (1982) also pointed to the importance of some forms of conflict in development by exploring the outcomes of nonconserving children's interactions on string of different conservation tasks. Further work has demonstrated that the benefits of interaction extend to promoting development in a range of different domains including science learning (Howe et al., 1992), moral reasoning (Damon & Killen, 1982; Leman & Duveen, 1999), and inference tasks (Roazzi & Bryant, 1998; Tudge, 2000).

Our third hypothesis therefore predicted that conversation (interaction) between children on this (punishment) task will be associated with developmental changes in preference for different sorts of punishment. In other words, children who initially chose expiatory punishments and who engaged in conversation would be more likely to choose reciprocal punishments after interaction compared with those who did not engage in interaction. This prediction is consistent with the more general observation that social interaction between peers promotes development.

Additionally, for our fourth hypothesis, we predicted that these advances would not differ in terms of gender or in terms of the gender mix of the children involved in a discussion. This prediction was made on the basis of previous research that finds no link between gender and decision making or developmental change (Leman & Duveen, 1999; Leman et al., 2005). However, while we predicted no gender differences in development, there is some other research pointing to an influence of gender on development: On a conservation of liquids task, Psaltis and Duveen (2007) found that conversation "types" differ according to gender and that these, in turn, were associated with variations in development. The present study offers an opportunity to examine effects of gender on development in a context of moral judgments, which is a domain in which gender differences in moral reasoning and judgment have often been proposed.

Development and the Features of Conversation

Although the association between interaction and learning is well established, the matter of identifying precisely which conversational features promote learning and development has proved surprisingly elusive. In a series of systematic studies exploring children's discussions about a range of group-based, science tasks, Howe found no consistent links between the quality or quantity of arguments that children produced in interaction and subsequent development (e.g., Howe, McWilliam, & Cross, 2005; Howe, Tolmie, Duchak-Tanner, & Rattray, 2000; Howe et al., 1992). Consequently, Howe proposed an *incubation hypothesis* whereby interaction between peers promotes development by stimulating consideration of alternative explanations for a particular scientific phenomenon. In short, Howe's hypothesis is that interaction between peers primes a child to search, individually, for explanations for a phenomenon after the period of interaction has finished. Thus the developmental "work" takes place on an individual level and is not a feature of group discussion itself. The incubation hypothesis provides an explanation why researchers have had such difficulty linking specific features of interaction with development because the benefits of peer interaction are not, necessarily, related to the quality (or quantity) of argument and explanation in peer interaction. Howe et al. (2005) argue that peer discussion serves an important role because it primes the search for information, and points children to the possibility of alternative explanations.

One might anticipate that gender influences on conversation dynamics will influence development too. Much research demonstrates that boys' dominance of classroom resources can set up marked differences in the opportunities for learning for boys and girls (e.g., Littleton, Light, Joiner, Messer, & Barnes, 1998; Strough, Berg, & Meegan, 2001). Yet surprisingly few studies have sought to establish whether any such gender differences feed into longer term learning, and those that have provided somewhat contradictory results (Jones et al., 2000; Psaltis & Duveen, 2006).

Our fifth hypothesis, following Howe et al. (2005), was that there would be no relation between conversation content (the arguments or justifications that children produce in conversation) and developmental changes in judgments about punishments. In other words, no particular argument or set of arguments would be associated with developmental changes. Given ambiguous previous research findings, we did not propose a hypothesis regarding the links between conversation dynamics and development.

Method

Design

The study was in four phases. In the first phase, all participants independently judged which was the fairest of two punishments in each of a set of stories concerning a child's misbehavior (see the Materials section, below). In the second phase (which occurred on the same day as the first phase), those children who were not in the control group were placed in a pair with a peer who had given a different response to one of the stories, *The Broken Window*, and were asked to discuss this and to reach an agreement about what the fairest punishment was. Children in the control (no interaction) group were asked to give a response on their own. The time interval between first and second test phases was kept roughly the same for all children, between 1 and 2 hr.

Two further, postinteraction tests were conducted. The third phase was conducted 2 weeks after the initial session and all children were asked, again independently, their response to the broken window story. Finally, for the fourth phase 8 weeks after the interaction phase, all participants were asked to make judgments on all stories once again, independently.

Pairs comprised children from the same year group at school who were identified by the teacher as neutral in terms of friendship; that is, they were neither good friends nor enemies. Otherwise, pairs were selected at random although, because a key aim was to establish the effects of gender on conversation and judgments, gender was used as a blocking variable. Thus, there was one key independent variable, pair type, that varied by the gender mix of the pair and children's initial responses (see Table 1) with a control group of children who did not engage in discussion with a peer on the task.

There were several dependent variables. There were two categories of outcome variable (moral judgment decisions). First, there were the children's responses to the stimulus story about the broken window that were either reciprocal or expiatory. These responses were given at four time points; *t*1 (preinteraction), *t*2 (either immediately postinteraction or 1 to 2 hr later), *t*3 (2 weeks later), and *t*4 (8 weeks after the initial session, *t*1). Second, there was a measure of punishment orientation that was based on children's responses to three punishment stories presented at *t*1 and *t*4. (Initially, four stories were presented to children. However, on reflection, it was felt that one of the stories might appear to be unusual for children in terms of normal adult–child relations, and so this story was not included in the calculation of the global measure of judgment at *t*1 and *t*4.) As each story yielded either a reciprocal or expiatory response, scores were added to create a total score out of three which represented the maximum reciprocal score based on the three stories. Finally, there were two further sets of conversation measures for children involved in interaction: (a) children's use of simultaneous speech acts (overlaps, negative and positive interruptions) and (b) types of justifications given.

Participants

A total of 133 children participated in the study. Children came from two consecutive year (age) groups in the same school in Berkshire, Southern England. In one year group, there were 42 boys and 26 girls (8.5–9.5 years; *M* = 9.10, *SD* = 0.71). In the other year group were 35 boys and 30 girls (9.4–10.5 years; *M* = 9.92, *SD* = 0.66). The average age of the sample as a whole was 9.58 years.

The mix of ethnicities in the school was broadly homogenous (96% White British) and they were all from middle-class backgrounds. Children were recruited through letters home to their parents or guardians, who were required to give written consent for their child's participation before the initial data collection phase. The children were also required to consent verbally to their participation at all stages.

Materials

Stimulus stories were selected after pretesting materials with five children of the same age as the participants in the main study. These children were read five stories, based on those originally developed by Piaget (1932, pp. 200–209), in order to determine whether they would invoke sufficient numbers of differing responses to make discussions possible. Each story provided a choice between two options which corresponded to either an expiatory punishment or a reciprocal punishment.

Following the pilot study, stories were chosen to be presented to children in the main study. In each story, children were told a brief description of an event involving a child who committed a moral transgression. Then, in each case, the child had to make a choice between alternative punishments

Table 1
Definition of Pair Types

Independent response				
Reciprocal	Expiatory	Pair type	Description	*N*
Girl	Girl	Girl–girl	Reciprocal female with expiatory female	24
Girl	Boy	Girl–boy	Reciprocal female with expiatory male	24
Boy	Boy	Boy–boy	Reciprocal male with expiatory male	24
Boy	Girl	Boy–girl	Reciprocal male with expiatory female	24
Girl–boy	Boy–girl	Control	Did not engage in interaction	37

that corresponded to more mature (reciprocal) and less mature (expiatory) answers.

The broken window story was used as the stimulus story for the children to discuss together in a pair because it had the closest to a 50% response rate for each of the alternatives (reciprocal and expiatory). The text of the story is as follows:

> Late one afternoon there was a boy who was playing with a ball on his own in the garden. His dad saw him playing with it and asked him not to play with it so near to the house because it might break a window. The boy didn't really listen to his dad, and carried on playing near the house. Then suddenly, the ball bounced up high and broke the window in the boy's room. His dad heard the noise and came out to see what had happened. The father wonders what would be the fairest way to punish the boy. He thinks of two punishments. The first is to say: "Now, you didn't do as I asked. You will have to pay for the window to be mended, and I am going to take the money from your pocket money." The second is to say: "Now, you didn't do as I asked. As a punishment you have to go to your room and stay there for the rest of the evening." Which of these punishments do you think is the fairest?

The story text was read out to the children and they were all shown a picture of the act (i.e., Sam breaking the window), along with pictures depicting the two choices of punishment. The text of the other stories is given in the Appendix.

In the interaction phase of the study, children discussed their differing responses to the first story about the boy who broke the window in pairs in order to together arrive at an agreement on which was the preferable punishment. Children's conversations were videotaped using a camera and specialist microphone.

Procedure

In the independent response parts of the study (i.e., Phases 1, 3, and 4), the participating children were presented a booklet in a paper-and-pencil test of moral judgment, which was carried out in groups of 15–20 children. Each story was read out to the children and alongside this the child was shown a picture of the event or misdeed that lead to the punishment. The two punishments were then read out to the child who was also shown the corresponding pictures alongside one another. The child was asked to indicate, by ticking the relevant picture, which response he or she felt was the fairest. This procedure was then repeated with all four stories at *t*1 and *t*4. For children at *t*2 and *t*3 only material relating to the broken window story was presented.

For the second part of the study children were paired together with a partner who in the preinteraction phase had given a contrasting response to the broken window story. In addition, pairs comprised one child who was classified as reciprocal *overall* (three out of four responses to the stories had been reciprocal) and one child who was classified as expiatory overall (three out of four expiatory).

Children were allocated at random to control or interaction groups. Those children who were selected to participate in interaction were placed in pairs on the basis of their sex (with sex used as a blocking variable). The control group of children did not participate in the interaction phase. Pairs were seated in a quiet room and were read the story. They were reminded of their individual responses given in the first phase and were then asked to discuss the task together and to try to reach agreement. During the discussion, the interviewer left the room so the children could talk undisturbed. In cases when the interviewer returned before consensus was reached, the children were asked to continue their discussion. All pairs reached an agreement after 5 min.

Analysis of Conversations

Conversation dynamics. Three categories of *simultaneous speech act* were examined in the study: *overlaps*, *positive interruptions*, and *negative interruptions*. Simultaneous speech acts occur in conversations when the next speaker starts speaking at a point in the current speaker's turn that would not be defined as being at or close to a transition relevant place (Jefferson, 1973). In the present study, an overlap was defined as simultaneous speech that constituted no bid to take the floor. This included back channel utterances and simultaneously uttering the same word or phrase. It also covered cases when one speaker thought the other speaker had finished their turn, started speaking, but immediately stopped as they noticed that the partner was not yet done. Positive interruptions were defined as bids to take the floor that were not performed in a hostile or negative manner. These also included interruptions to make clarifications, reassertions of one's own position, or interruptions made to help the other finish a sentence. The latter would be a sign of active listenership that was performed in a somewhat collaborative manner, as the speakers could be perceived as constructing the utterance together. Negative interruptions were interruptions where the speaker attempted to take the floor in a hostile fashion. These included successful interruptions such as when the interrupter succeeds in taking the turn, and interruptions made in an aggressive or dominant manner. A count of each different type of interruption was made which, necessarily, occurred at turn boundaries.

Conversation content. We were also interested in examining the justifications used by the children to convince a conversation partner. To identify an appropriate list of such justifications, we employed an iterative qualitative process that commenced with listing *all* justifications used throughout the conversations. The next stage was to examine this list of justifications to identify common themes or categories of justification. Through this iterative process these justifications were refined into seven categories that covered all justifications observed throughout the corpus of conversational data and that were relevant to the study (see Table 2).

All justifications observed in the study were assigned to a category and the total number of justifications used within each category was counted. A child could use more than one justification and could include more than one type of justification in each turn. The total use of each justification category by each child was counted, as well as the proportionate use of each category. The proportionate use was calculated dividing all uses of one justification category with the total uses of justifications by one child.

Interrater reliability. Conversations were coded by the second author. For simultaneous speech acts and justification categories, 15 conversations (about 30% of the total number of conversations) were also coded by the first author. Interjudge agreement for simultaneous speech acts was excellent (kappa = .97). There was also excellent agreement between judges for the justification categories (kappa = .88).

Results

Statistical Note

As the focus of this study is dyadic interactions the issue of nonindependence needs to be accounted for. In conversation with their partner the participants are likely to adapt their behavior according to the behavior of the other. As has been pointed out (Carli, 1989a, 1989b) in a study where the gender mix of the pair varies, it is likely that the participants change their conversation and interaction according to both their own gender and the gender of their conversational partner. In the present study, this issue of nonindependence has been addressed using the same approach as Leman et al. (2005). Separate analyses of variance (ANOVAs) for the data for same- and mixed-gender pairs were performed. In the first analysis, gender differences were compared between boys in the all-boy pairs and girls in the all-girl pairs. In the second analysis, the comparison was between the boys

Table 2
Justification Categories With Corresponding Examples

Category	Definition	Example
Assertion	Assertions that only state the subject's belief, without any justifications	"I think staying in his bedroom"
Intention	Reference to the intention of the transgressor	"It might have been an accident" "He didn't mean to do it"
Reference to authority	Reference to what is right according to an authority figure	"My mum and dad did that to me"
Reciprocity	Putting right what is wrong	"He broke it and it's only fair that what he breaks he will replace I think"
Moral	Reference to a moral rule	"He didn't do as he was told"
Expiatory force	The punishment is preferred because it is severe	"If he just went to his room he might think that he can do it again because that's not a hard punishment"
Avoiding expiation	The punishment is preferred because it is mild, or the alternative is too severe.	"It should be stay in his room because then he can play"

Table 3

Mean Number of Overlaps, Positive Interruptions, and Negative Interruptions for Boys and Girls in Same- and Mixed-Gender Pairs and Associated Univariate Effects

	Same-sex pairs						Mixed-sex pairs					
	Boys $N = 24$		Girls $N = 24$				Boys $N = 23$		Girls $N = 23$			
Variable	*M*	*SD*	*M*	*SD*	*F*(1, 46)	Partial η^2	*M*	*SD*	*M*	*SD*	*F*(1, 44)	Partial η^2
Overlaps	0.83	1.01	0.67	0.96	0.343	.007	0.83	1.30	0.52	0.79	0.92	.020
Positive	0.50	1.18	0.25	0.44	0.945	.020	0.17	0.39	0.22	0.52	0.10	.002
Negative	0.63	1.06	0.21	0.41	3.24	.066	0.04	0.21	0.04	0.21	0.00	.000

and the girls in mixed-gender pairs. The means and mean error scores are then used in a set of formulae that yield separate *t* statistics concerning gender. The different comparisons all have the following denominator: $[(2\mathrm{MS}_e + 2\mathrm{MS}'_e)(1/n)]^{1/2}$, in which MS_e is the mean error score for gender effects in same-gender pairs, MS'_e is the mean error score for gender effects in mixed-gender pairs and n is the number of observations on which the means were calculated. The numerator for speaker gender effects is $(M_{gs} + M_{gm} - M_{bs} - M_{bm})$ and for partner gender effects $(M_{gs} - M_{gm} + M_{bm} - M_{bs})$, where M_{gs} is the mean for girls in same-gender pairs, M_{gm} is the mean for girls in mixed-gender pairs, M_{bs} is the mean for boys in same-gender pairs, and M_{bm} is the mean for boys in mixed-gender pairs. The interaction effect contrasts any gender differences found in same- and mixed-gender pairs to investigate in which pair type greater gender differences can be observed. The interaction effect has the numerator $(M_{gs} - M_{gm} - M_{bm} + M_{bs})$. For the speaker and partner gender effects a negative value of *t* indicates a higher score for boys than girls. Regarding the interaction effects, a negative value of *t* demonstrates that a greater difference can be observed in the mixed-gender compared to the same-gender pairs.

Analyses of Conversations

Conversation dynamics. Simultaneous speech acts were analyzed using the correction for nonindependence. Mean values for overlaps, positive interruptions, and negative interruptions, by gender and pair type, are given in Table 3, which also includes the results from the two ANOVAs: gender differences between boys and girls in mixed-gender pairs, and gender differences between boys and girls in same gender (i.e., all boy versus all girl) pairs. In Table 4, the *t* values and significance tests for children's use of simultaneous speech acts are presented.

Boys used more overlapping speech than girls. There were no significant differences regarding partner gender or interaction effects. There were no significant differences attributable to the gender of the speaker in terms of positive interruptions. However, all participants used more positive interruptions when talking with a boy than with a girl. An interaction effect was also observed, indicating that there was a greater gender difference in the same-gender pairs than in the mixed-gender pairs; both boys and girls use less positive interruptions when talking to someone of the opposite sex, with a greater decrease in the boys' use of positive interruptions when talking to a girl.

Boys used more negative interruptions than girls, and all children used more negative interruptions when paired with a boy. There were greater gender differences in the same-gender pairs than in the mixed-gender pairs. This difference was due to both boys and girls producing fewer negative interruptions in mixed-sex conversations, with a larger

Table 4

Speaker Gender, Partner Gender, and Interaction Effects Relating to Simultaneous Speech Acts

Variable	Speaker gender *t*(92)	Partner gender *t*(92)	Interaction *t*(92)
Overlaps	−2.25*	0.66	0.73
Positive interruptions	−1.22	−1.73*	2.12*
Negative interruptions	−2.84**	−2.84**	5.09***

Note. For speaker and partner effects, a negative value of *t* indicates a higher score for boys than girls. A negative value of *t* in the interaction effect indicates that there was a greater gender difference in mixed-gender compared to same-gender pairs.

*$p < .05$. **$p < .01$. ***$p < .001$.

Table 5
Mean Use of Justifications for Boys and Girls in Same- and Mixed-Gender Pairs and Associated Univariate Effects

	Same-gender pairs						Mixed-gender pairs					
	Boys $N = 24$		Girls $N = 24$				Boys $N = 23$		Girls $N = 23$			
	M	*SD*	*M*	*SD*	$F(1, 46)$	Partial η^2	*M*	*SD*	*M*	*SD*	$F(1, 44)$	Partial η^2
Justifications												
Assertions	1.00	0.98	0.29	0.46	10.27	.183	0.70	0.82	0.35	0.71	2.35	.051
Reciprocity	0.13	0.34	0	0	3.29	.067	0.13	0.34	0.09	0.29	0.22	.005
Intention	0.25	0.61	0.13	0.34	0.78	.017	0.17	0.39	0.30	0.56	0.85	.019
Authority	0.08	0.28	0.08	0.28	0	0	0.09	0.29	0	0	2.10	.045
Moral	0.08	0.28	0	0	2.09	.043	0.13	0.46	0.17	0.39	0.12	.003
Expiatory force	0.42	0.78	0.13	0.45	2.54	.052	0.22	0.52	0.17	0.49	0.09	.002
Avoiding expiation	1.00	1.32	0.92	1.06	0.06	.001	0.82	0.98	0.96	1.33	0.14	.003

decrease in the boys' use of negative interruptions in mixed-gender pairs.

Conversation content. Justification categories were analyzed using the correction for nonindependence. The results from the mixed-gender and same-gender ANOVAs, and the mean values for the children's use of the different justification categories by gender and by pair type are shown in Table 5. The *t* scores and significance tests for the justifications used by the children are given in Table 6.

No significant differences between boys and girls were found in the use of justifications related to the categories reciprocity, intention, authority, moral, and avoiding expiation for either speaker gender, partner gender, or interaction effects.

However, boys used significantly more assertion justifications (merely restating one's own position without further elaboration) than girls, and all participants used more assertion justifications when talking with a boy than with a girl. In terms of expiatory force justifications, boys used these types of justifications significantly more frequently than girls. However, no differences for partner gender effects or interaction effects were observed.

The conversations varied in length, so the proportionate use of the justification categories was also calculated. To correct for non-normal distribution, a square root transformation was conducted and the transformed data were then analyzed using the correction for nonindependence. Significant gender differences in the proportionate use of assertion justifications were observed, $t(92) = -2.97$, $p < .01$, with boys using these more than girls. No partner or interaction effects were observed, and there were no other gender differences relating to the proportionate use of any of the other justification categories.

Table 6
Speaker Gender, Partner Gender, and Interaction Effects Relating to the Justification Categories

Variable	Speaker gender $t(92)$	Partner gender $t(92)$	Interaction $t(92)$
Assertions	−5.84***	−1.99*	1.37
Reciprocity	−1.55	−0.75	−0.85
Intention	0.04	−1.78	−0.72
Authority	−0.85	0.85	0.78
Moral	−0.35	−1.10	−1.92
Expiatory force	−2.15*	−1.59	0.97
Avoiding expiation	0.21	−0.95	0.60

Note. For speaker and partner effects, a negative value of *t* indicates a higher score for boys than girls. A negative value of *t* in the interaction effect indicates that there was difference in mixed-gender compared to same-gender pairs.
*$p < .05$. ***$p < .001$.

Moral Judgments: Changes in Concepts of Punishment

Preinteraction test (t1). There were no gender differences in terms of responses to any of the stories at any stage. However, there was a significant association between age group and response on the broken window story at *t*1, $\chi^2(1) = 3.86$, $p < .05$, contingency coefficient = .20. In the younger age group, 43% gave reciprocal response compared with 65% in the older age group. There was also a significant association between age group and response on the taking turns story at *t*1, $\chi^2(1) = 3.97$, $p < .05$, contingency coefficient = .20.

The more mature, reciprocal response was given by 75% of the younger age group and 90% of the older age group, respectively.

An overall punishment orientation score was calculated by adding together the total number of reciprocal responses to each story (where 3 was the maximum, indicating all reciprocal responses, and 0 the minimum, indicating all expiatory responses). A two-way Age Group × Gender ANOVA revealed a main effect of age group, $F(1, 132) = 7.34$, $p < .01$, partial $\eta^2 = .054$, with older children ($M = 2.28$, $SD = 0.74$) giving more reciprocal answers than younger children ($M = 1.94$, $SD = 0.71$).

There were no differences between those who were selected for interaction and those who were not in terms of overall score. As all pairs comprised one reciprocal and one expiatory child, there were no differences by pair type. There was no significant difference comparing the responses of children in the control group with the interaction groups. This analysis of age group differences is broadly consistent with Piaget's (1932) prediction that judgments in this domain undergo qualitative changes (from expiatory to reciprocal orientations) at around 8–11 years.

Immediate postinteraction or short delay test (t2). For children involved in interaction, *t*2 was the joint (agreed) response of the pair. For children in the control group (not involved in interaction) the *t*2 was delayed at around 1 or 2 hr after *t*1, which was a similar amount of time to the delay for the children in interaction. This second time point, *t*2, concerned only responses on the broken window story. There were no statistically significant differences attributable to pair type (and consequently, to gender) in this immediate posttest, although as can be seen from Figure 1, in all but one pair there was a slight drop in the proportion of reciprocal (more advanced) responses.

Two-week delayed posttest (t3). A 2-week delayed posttest (*t*3) asked children individually once again for their responses only to the broken window story. Again, analysis revealed no significant differences between those who had been in a pair and those who had not in terms of their responses to the stimulus story. There were, again, no differences in terms of the pair gender type, or gender, in terms of responses to the story.

Eight-Week Delayed Posttests (t4). In the 8-week delayed posttest (*t*4), responses were again collected individually for all stories. Examining each story in turn, there were no significant differences in responses to any of the individual stories between those who had, and those who had not, been involved in interaction, and no gender differences in responses to individual stories.

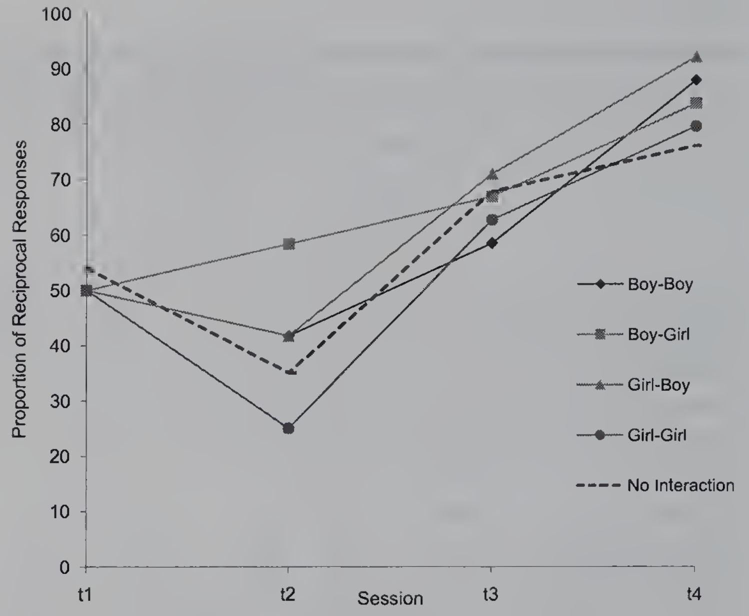


Figure 1. Proportion of reciprocal responses to the broken window story by pair type at *t*1, *t*2, *t*3, and *t*4.

However, overall (i.e., adding responses from three stories together) there was significant variation by pair type in punishment orientation scores, $F(4, 132) = 5.14$, $p < .001$, partial $\eta^2 = .138$ (see Figure 2). Post hoc Tukey honestly significant difference (HSD) tests ($p < .05$) revealed that the mean in the control group ($M = 2.27$, $SD = .73$) differed significantly from the mean in the boy–boy ($M = 2.79$, $SD = .42$), boy–girl ($M = 2.71$, $SD = .46$), and girl–boy ($M = 2.83$, $SD = .38$) pairs. The difference

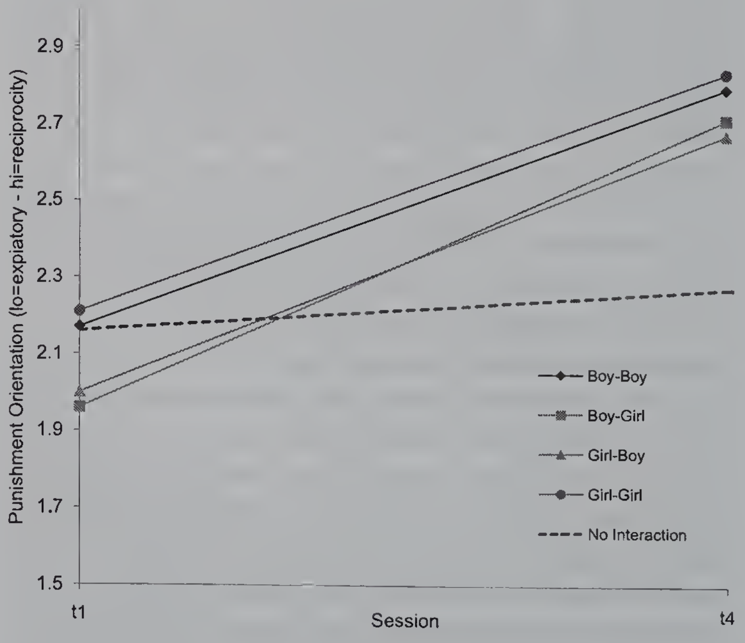


Figure 2. Mean punishment orientation score (0 = *all expiatory* to 3 = *all reciprocal*) by pair type at *t*1 and *t*4.

approached significance ($p = .061$) for the comparison between the control group and girl–girl ($M = 2.67$, $SD = .64$). Comparisons of boys and girls and of those only in the control group revealed no gender differences in orientation.

Changes in punishment orientation (t1–t4). To explore changes in punishment orientation, a repeated measures ANOVA compared responses by pair type at *t*1 and *t*4. This revealed a large and significant effect of the repeated measure, $F(1, 128) = 69.71$, $p < .001$, partial $\eta^2 = .353$. However, this was qualified by a significant effect of pair type, $F(4, 128) = 4.49$, $p < .01$, partial $\eta^2 = .123$. Post hoc follow-up *t* tests compared changes in punishment orientation for each pair type. These tests revealed higher scores (i.e., more reciprocal orientation) at 8 weeks (*t*4) than at initial testing (*t*1) for all pair types—boy–boy, $t(23) = 3.98$, $p < .001$; boy–girl, $t(23) = 4.10$, $p < .001$; girl–boy, $t(23) = 5.82$, $p < .001$; girl–girl, $t(23) = 3.82$, $p < .001$—with the exception of the control group, $t(36) = 0.85$, $p = .40$ (see Figure 2).

Conversation and moral development. In order to examine the impact of conversation features (content and dynamics) on outcomes, separate logistic regressions (forward stepwise, likelihood ratio method) were conducted for each of the test points (i.e., *t*2, *t*3, and *t*4). At *t*2 (i.e., leading up to the joint response of the pair) only the proportionate use of expiatory force justifications was found to relate to outcomes, $\chi^2(1) = 9.02$ $p < .001$, Wald = 6.40. The relation was positive, indicating that use of these justifications often led to a reciprocal joint reply. The analysis at time *t*3 revealed that the use of expiatory force justifications was once again positively related to outcomes, $\chi^2(1) = 6.02$, $p < .001$, Wald = 3.53. No features of conversation content or dynamics were associated with outcomes at the 8-week posttest (*t*4).

Discussion

The present study explored the relations among gender, conversation, and children's judgments about what constitutes a fair punishment. The results demonstrate that gender influences the dynamics of children's conversations. Moreover, conversation was associated with development: Children who discussed the task made more mature, reciprocal judgments than those who did not. Yet although gender affects conversation, and conversation leads to development, variations in the dynamics and content of conversations do not appear to carry over to influence learning and development.

Gender and Conversation

Our first hypothesis predicted that boys would use more simultaneous speech acts (overlaps and interruptions) than girls, and this was confirmed. Specifically, boys used more overlaps and negative interruptions than girls. In addition, all children used more interruptions with a boy. However, our second hypothesis was not confirmed. We predicted that gender differences would be accentuated in boy–girl pairs compared with the same-sex pairs. In fact, there was a greater difference when comparing boys and girls in same-gender (as opposed to mixed gender) pairs. Thus, in contrast to previous findings, in the present study boys and girls in mixed-sex interaction relied less on "male" and "female" conversational styles (at least in the use of simultaneous speech acts).

Results from the present study present us with partial confirmation of previous work on gender and conversation dynamics but also raise further questions about how children respond to mixed-sex interactions. The failure to confirm our second hypothesis might be attributable to several factors; for instance, compared with other studies there may have been differences in the way in which the coding scheme was applied or in sample characteristics. However, perhaps a key difference is that here we used a moral judgment story and although children in a pair had differing perspectives (expiatory and reciprocal responses) the tasks could be seen as relatively nonconflictual. In other studies (e.g., Leman et al., 2005) tasks may involve explicit conflict between children. Moreover, in important respects the present findings match work with preschoolers and older children (Killen & Naigles, 1995; Nelson-Le Gall & DeCooke, 1987), which found that gender differences were less marked in mixed-gender interaction compared with same-gender interaction, because in the former children used more collaboration whereas in the latter they used more commands. How and why different tasks might inspire different gender dynamics in interaction remains an open question. However, the results from the present study demonstrate the importance of context in eliciting different forms of interaction between children generally (Oliveira & Sadler, 2008), and in terms of gender dynamics specifically (Holmes-Lonergan, 2003). Future research is required to establish what features of context may interact with gender, and other social factors, to affect conversation dynamics.

There were gender differences in terms of the use of only two categories with regard to the content of children's conversations. Boys used more assertions than girls, and all children used more assertions when speaking with a boy. When children used an assertion they merely stated or restated a child's belief (e.g., ''He should stay in his room,'' ''Making him pay is fairer'') and did not offer any justification or explanation to support that belief. In conversational terms, assertions may not only communicate a factual statement of belief but also act as a kind of rhetorical gambit: Stating one's position or doing so repeatedly or in the face of another's contrary belief may also constitute an attempt to set or dominate the conversational agenda. Thus, the present finding that boys used more assertion in conversation than girls may reflect more general gender differences in conversation dynamics or style rather than any unwillingness or inability of the boys to elaborate on their positions.

Gender and Moral Judgment

There was also an effect of speaker gender on the use of justifications that appealed to *expiatory force*. These justifications point to the severity of punishment (i.e., a punishment is deemed fair because it is severe) and are closely associated with the expiatory position. In the present study, boys used more of these sorts of justifications than girls although equivalent numbers of boys and girls made expiratory judgments. It may be the case that the expiatory boys were more vociferous in arguing their position than expiatory girls in the present study.

Previous research suggests that there are few, if any, reliable gender differences in moral orientation (e.g., Walker, 2006). One might anticipate that differences in orientation would become evident in children's conversations, especially when children were given the task of justifying their own perspective that contrasted with that of a peer. But the results from the present study indicate that there were few gender differences in the ways in which children justified their judgments. This suggests that any underlying orientations to morality do not divide simply along gender lines at this age, although gender certainly is an important factor affecting the dynamics of children's interactions. Thus, it was not the case that the children were unaware of or insensitive to gender as a social category, or that boys and girls at this age do not reproduce typical ''male'' and ''female'' conversational styles. However, these effects of gender are not visible in terms of moral judgments and justifications.

Conversation and Development

Our third hypothesis predicted that interaction between children on this task would be associated with changes in punishment orientation (compared with a no-interaction, control group of children). Additionally, our fourth hypothesis predicted that there would be no relation between gender, or the gender mix of children involved in interaction, and developmental advances. Both of these hypotheses were confirmed.

On one hand, it should not come as a surprise that children who engaged in interaction with a peer changed their judgments to the more adaptive, reciprocal orientation. Previous research has demonstrated the benefits of interaction for the development of moral judgment (Damon & Killen, 1982; Leman & Duveen, 1999). Moreover, our finding that the gender of children engaged in conversation had no influence on developmental outcomes is also consistent with previous work on this topic (Leman et al., 2005). On the other hand, it is striking that such a brief period of conversation between children on a single vignette should have an influence on moral judgment after 8 weeks. The finding is striking, but it is consistent with other work (e.g., Psaltis & Duveen, 2006, 2007). Moreover, it is plausible to argue that once the ''penny drops'' in terms of an individual's understanding it is unlikely that the change in the basis of moral judgments will be quickly undone.

It is important, however, to include a note of caution in terms of limitations to the tasks that we used in the present study. In many of Piaget's (1932) vignettes the role of authority is often confounded with other significant aspects of a task that may obscure the nature of children's understanding (Costanzo, Coie, Grumet, & Farnhill, 1973). In this respect our punishment stories, like Piaget's, present children with situations where parents have the power to punish children and hence inevitably incorporate parental authority as an ingredient in the judgment that participants were asked to make. Although our observations during pretesting and data collection suggest that children did not have difficulty imagining that adults might choose between expiatory and reciprocal punishments, a limitation of the vignettes we used is that children may have responded in ways that sought to anticipate how an adult authority figure would behave rather than select the fairest punishment. Caution is therefore required in extending research findings to *all* children's interactions and, in particular, peer interactions which differ in important ways from adult–child interaction.

A further important qualification is that the confirmation of our third and fourth hypotheses was possible for our global (or overall) measure of punishment orientation, but not for judgments about the conversation stimulus story (the broken window). This suggests that any developmental benefits of conversation extend across a subdomain (conceptions of punishment) rather than being restricted to a particular story. In other words, this is development at a deep, structural level rather than merely a learned response to a particular story.

Development and the Features of Conversation

The final focus for the present study was on the links between features of conversations (content, dynamics, or interruptions) and developmental change. Ultimately, we found no relation between conversation content and development, although in the shorter term, justifications pointing to the use of expiatory force did relate to outcomes. From an educational and developmental perspective, the important finding was that features of conversations did not relate to judgments at 8 weeks after interaction, which was consistent with our fifth hypothesis.

The present findings are consistent with previous work (notably Howe et al., 2005) that has found no links between specific features of conversations and development. However, our findings hint that explicitly discussing justifications involving expiatory force may promote perspective taking in the short- to midterm. In this respect, it is important to note that justifications that focus on expiatory force are closely aligned to the *less* mature moral response. It may be that explicitly examining, and then rejecting, the justifications for a less mature response, helps children to appreciate the legitimacy of the more advanced position more quickly. Such an interpretation is certainly also consistent with Psaltis and Duveen's (2007) study, which indicated that explicit recognition of the reasons or grounds for an answer was closely associated with development after interaction.

We also found no links between conversation dynamics (gauged here by use of simultaneous speech acts) and (developmental) outcomes. The failure to find any connection between conversation dynamics and outcomes indicates that although gender (and the gender mix of children involved in interaction) affects conversation dynamics, this does not necessarily feed through to influence learning and development, or even shorter term decision making. At this age, children can ''work around'' the effects of gender on conversation in terms of decision making and learning (Leman & Lam, 2008; Leman et al., 2005); that is, although boys' and girls' conversations show distinctive characteristics, and although boys and girls react to the gender of their conversation partner in terms of their conversation styles, these do not ultimately affect learning. However, it should not be forgotten that gender effects on conversation may be insidious: For instance, boys' dominance may reinforce status hierarchies and differences between genders from an early age. Moreover, conversation and interaction (in the classroom and elsewhere) provide a key forum through which stereotypes may be communicated.

Conclusions. These findings demonstrate that conversation between peers leads to development. Moreover, there were no effects of conversation dynamics and conversation content on longer term development. Thus, conversation on a topic may stimulate a process of individual reflection that triggers developmental advances. However, it seems that the process may be accelerated by explicit recognition, in a conversation, of the reasons against a less advanced position or, to follow Psaltis and Duveen (2006, 2007), the reasons for a more advanced position. An implication of this finding is that educators do not necessarily need to guide the contents of a specific conversation between children too closely. Rather, following Howe et al. (2005), learning benefits are a consequence of postconversation reflection.

The present findings also suggest that educators, policy makers, and others need to be cautious in jumping to the conclusion that because gender can influence the dynamics of interaction in problem-solving tasks and the classroom it is inevitably detrimental to the learning of one or both gender groups. By 7 years, children are skilled social actors who recognize the influence of gender and other social categories such as race on social relations and interpersonal behavior (Leaper & Smith, 2004; Leman & Lam, 2008). In some contexts at least, children can work around these influences to make good judgments regardless of the dynamics or style of a conversation.

References

Ames, G. J., & Murray, F. B. (1982). When two wrongs make a right: Promoting cognitive change by social conflict. *Developmental Psychology, 18*, 894–897.

Brinker, S. R., Goldstein, S. E., & Tisak, M. S. (2003). Children's judgments about common classroom punishments. *Educational Research, 45*, 189–198.

Carli, L. L. (1989a). Gender differences in interaction style and influence. *Journal of Personality and Social Psychology, 56*, 565–576.

Carli, L. L. (1989b). Gender differences in interaction style and influence: Correction. *Journal of Personality and Social Psychology, 57*, 964.

Coates, J. (1989). Gossip revisited: Language in all-female groups. In J. Coates & D. Cameron (Eds.), *Women in their speech communities* (pp. 94–122). New York: Longman.

Costanzo, P. R., Coie, J. D., Grumet, J. D., & Farnhill, D. A. (1973). A re-examination of the effects of intent and consequences on children's moral judgments. *Child Development, 44*, 154–161.

Damon, W., & Killen, M. (1982). Peer interaction and the process of change in children's moral reasoning. *Merrill-Palmer Quarterly, 28*, 347–367.

Doise, W., & Mugny, G. (1984). *The social development of the intellect*. Oxford, UK: Pergammon.

Doise, W., Mugny, G., & Perret-Clermont, A.-N. (1975). Social interaction and the development of cognitive operations. *European Journal of Social Psychology, 5*, 367–383.

Felton, M., & Kuhn, D. (2001). The development of argumentative discourse skills. *Discourse Processes, 32*, 135–153.

Fernandez-Berrocal, P., & Santamaria, C. (2006). Mental models in social interaction. *Journal of Experimental Education, 74*, 229–248.

Gilligan, C. (1982). *In a different voice: Psychological theory and women's development*. Cambridge, MA: Harvard University Press.

Holmes-Lonergan, H. A. (2003). Preschool children's collaborative problem-solving interactions: The role of gender, pair type and task. *Sex Roles, 48*, 505–517.

Howe, C., McWilliam, D., & Cross, G. (2005). Chance favours only the prepared mind: Incubation and the delayed effects of peer collaboration. *British Journal of Psychology, 96*, 67–93.

Howe, C., Tolmie, A., Duchak-Tanner, V., & Rattray, C. (2000). Hypothesis-testing in science: Group consensus and the acquisition of conceptual and procedural knowledge. *Learning and Instruction, 10*, 361–391.

Howe, C., Tolmie, A., & Rodgers, C. (1992). The acquisition of conceptual knowledge in science by primary school children: Group interaction and the understanding of motion down an inclined plane. *British Journal of Developmental Psychology, 10*, 113–130.

Jaffee, S., & Hyde, J. S. (2000). Gender differences in moral orientation: A meta-analysis. *Psychological Bulletin, 126*, 703–726.

Jefferson, G. (1973). A case of precision timing in ordinary conversation: Overlapped tag-positioned address terms in closing sequences. *Semiotica, 9*, 47–96.

Jones, M. G., Brader-Araje, L., Wilson-Carboni, L., Carter, G., Rua, M. J., Banilower, E., et al. (2000). Tool time: Gender and students' use of tools, control, and authority. *Journal of Research in Science Teaching, 37*, 760.

Killen, M., & Naigles, L. R. (1995). Preschool children pay attention to their addressees: Effects of gender composition on peer disputes. *Discourse Processes, 19*, 329–345.

Leaper, C., & Smith, T. E. (2004). A meta-analytic review of gender variations in children's language use: Talkativeness, affiliative speech, and assertive speech. *Developmental Psychology, 40*, 993–1027.

Leaper, C., Tenenbaum, H. R., & Shaffer, T. G. (1999). Gender effects on the communication strategies of African American children from low-income, urban backgrounds. *Child Development, 70*, 1489–1503.

Leman, P. J., Ahmed, S., & Ozarow, L. (2005). Gender, gender relations and the social dynamics of children's conversations. *Developmental Psychology, 41*, 64–74.

Leman, P. J., & Duveen, G. (1999). Representations of authority and children's moral reasoning. *European Journal of Social Psychology, 29*, 557–575.

Leman, P. J., & Lam, V. L. (2008). The influence of race and gender on children's conversations and playmate choices. *Child Development, 79*, 1330–1344.

Littleton, K., Light, P., Joiner, R., Messer, D., & Barnes, P. (1998). Gender, task scenarios and children's computer-based problem solving. *Educational Psychology, 18*, 327–340.

Nelson-Le Gall, S., & DeCooke, P. A. (1987). Same-sex and cross-sex help exchanges in the classroom. *Journal of Educational Psychology, 79*, 67–71.

Oliveira, A. W., & Sadler, T. D. (2008). Interactive patterns and conceptual convergence during student collaborations in science. *Journal of Research in Science Teaching, 45*, 634–658.

Piaget, J. (1932). *The moral judgment of the child*. London: Routledge & Kegan Paul.

Psaltis, C., & Duveen, G. (2006). Social relations and cognitive development: The influence of conversation type and representations of gender. *European Journal of Social Psychology, 36*, 407–430.

Psaltis, C., & Duveen, G. (2007). Conservation and conversation types: Forms of recognition and cognitive development. *British Journal of Developmental Psychology, 25*, 79–102.

Roazzi, A., & Bryant, P. (1998). The effects of symmetrical and asymmetrical social interaction on children's logical inferences. *British Journal of Developmental Psychology, 16*, 175–181.

Saunders, B. J., & Goddard, C. (2007). The importance of listening to children: A qualitative study on the use of parental physical punishment in childhood in Australia. *Social Development Issues, 29*, 33–46.

Strough, J., Berg, C. A., & Meegan, S. P. (2001). Friendship and gender differences in task and social interpretations of peer collaborative problem solving. *Social Development, 10*, 1–22.

Tisak, M. S. (1993). Preschool children's judgments of moral and personal events involving physical harm and property damage. *Merrill-Palmer Quarterly, 39*, 375–390.

Tudge, J. (2000). Theory, method, and analysis in research on the relations between peer collaboration and cognitive development. *Journal of Experimental Education, 69*, 98–112.

Walker, L. J. (1984). Sex differences in the development of moral reasoning: A critical review. *Child Development, 55*, 677–691.

Walker, L. J. (2006). Gender and morality. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 93–115). Mahwah, NJ: Erlbaum.

Appendix: Additional Punishment Stories

Taking Turns

A boy was playing on a slide in the garden with his little brother, and wouldn't let his brother have a turn. The little brother got very upset and started to cry. The boys' dad saw this and decided he needed to punish the older boy. He thinks of two punishments. The first is to say: ''You wouldn't let your little brother have a turn. Now, I'm going to let your little brother have two turns, and you will miss a turn.'' The second is to say: ''You wouldn't let your little brother have a turn. Now, I'm going to take away one of your favourite toys until you learn to play properly.'' Which of these punishments do you think is the fairest?

The Broken Doll

A girl was playing with her little sister's things and wasn't being careful. She broke her little sister's favourite doll. Her mum finds out what has happened and thinks of two ways to punish her. The first is to say to the girl: ''You broke your sister's doll, so now you must go and buy her another one from your own money.'' The second is to say: ''You broke your sister's doll, so now you are not allowed to go and play at your friend's house.'' Which of these punishments do you think is the fairest?

Note. In each story, the first choice of punishment corresponds to the reciprocal orientation, the second to the expiatory orientation.

Child Development, May/June 2010, Volume 81, Number 3, Pages 972–987

Family Socioeconomic Status and Consistent Environmental Stimulation in Early Childhood

Robert Crosnoe
University of Texas at Austin

Tama Leventhal
Tufts University

R. J. Wirth
University of North Carolina at Chapel Hill

Kim M. Pierce
University of California, Irvine

Robert C. Pianta
University of Virginia

NICHD Early Child Care
Research Network

The transition into school occurs at the intersection of multiple environmental settings. This study applied growth curve modeling to a sample of 1,364 American children, followed from birth through age 6, who had been categorized by their exposure to cognitive stimulation at home and in preschool child care and 1st-grade classrooms. Of special interest was the unique and combined contribution to early learning of these 3 settings. Net of socioeconomic selection into different settings, children had higher math achievement when they were consistently stimulated in all 3, and they had higher reading achievement when consistently stimulated at home and in child care. The observed benefits of consistent environmental stimulation tended to be more pronounced for low-income children.

Ecological perspectives, which emphasize connections among multiple settings of children's lives, are highly relevant to education (Bronfenbrenner, 1979; Elder, 1998). For example, efforts to elucidate the ways in which families and schools come together to create unique spaces for children's development of cognitive and academic skills have informed not only theory but also policy, including No Child Left Behind (Epstein, 2005). This appreciation of multiple settings of development should be expanded to recognize the ways in which experiences in child care come together with experiences at home and in school to influence children's transitions into elementary school. Yet, any associations of proximal experiences in these settings with early learning must be parsed in relation to the larger

---

The National Institute of Child Health and Human Development (NICHD) Study of Early Child Care and Youth Development is directed by a Steering Committee and supported by NICHD through a cooperative agreement (U10), which calls for scientific collaboration between the grantees and the NICHD staff. The content is solely the responsibility of the named authors and does not necessarily represent the official views of the EKS National Institute of Child Health and Human Development, the National Institutes of Health, or individual members of the Network. Current members of the Steering Committee of the NICHD Early Child Care Research Network, listed in alphabetical order, are: Jay Belsky (Birkbeck University of London), Cathryn Booth-LaForce (University of Washington), Robert H. Bradley (University of Arkansas at Little Rock), Celia A. Brownell (University of Pittsburgh), Margaret Burchinal (University of North Carolina at Chapel Hill), Susan B. Campbell (University of Pittsburgh), Elizabeth Cauffman (University of California, Irvine), Alison Clarke-Stewart (University of California, Irvine), Martha Cox (University of North Carolina at Chapel Hill), Robert Crosnoe (University of Texas at Austin), James A. Griffin (NICHD Project Scientist and Scientific Coordinator), Bonnie Halpern-Felsher (University of California, San Francisco), Willard Hartup (University of Minnesota), Kathryn Hirsh-Pasek (Temple University), Daniel Keating (University of Michigan), Bonnie Knoke (RTI International), Tama Leventhal (Tufts University), Kathleen McCartney (Harvard University), Vonnie C. McLoyd (University of North Carolina at Chapel Hill), Fred Morrison (University of Michigan), Philip Nader (University of California, San Diego), Marion O'Brien (University of North Carolina at Greensboro), Margaret Tresch Owen (University of Texas at Dallas), Ross Parke (University of California, Riverside), Robert Pianta (University of Virginia), Kim M. Pierce (University of Wisconsin–Madison), A. Vijaya Rao (RTI International), Glenn I. Roisman (University of Illinois at Urbana-Champaign), Susan Spieker (University of Washington), Laurence Steinberg (Temple University), Elizabeth Susman (Pennsylvania State University), Deborah Lowe Vandell (University of California, Irvine), and Marsha Weinraub (Temple University).

Correspondence concerning this article should be addressed to Robert Crosnoe, Department of Sociology and Population Research Center, University of Texas at Austin, 1 University Station A1700, Austin, TX 78712-1088. Electronic mail may be sent to crosnoe@austin.utexas.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8103-0022

system of socioeconomic stratification in the United States. The need to do so is driven by the possibility of spuriousness in observed environmental effects on children (Duncan, Magnuson, & Ludwig, 2004), the important role of the transition into school in the intergenerational transmission of socioeconomic inequality (Brooks-Gunn & Duncan, 1997), and the long tradition of viewing both family–school connections and preschool as policy levers to reduce socioeconomic (and related) disparities among children (Epstein et al., 2002; Winsler et al., 2008).

This study, therefore, traces the reading and math trajectories of children who have experienced consistency or inconsistency in cognitive stimulation across family, child care, and classrooms settings during their entry into formal education. This investigation will also address the potential for the number and type of early sources of stimulation to contribute to socioeconomic disparities in learning as well as assess the potential for the socioeconomic advantages of parents (e.g., income, education, marital status, neighborhood composition) to explain and/or condition any observed effects of various configurations of early stimulation. These aims will be pursued by applying growth curve techniques to multisetting data from the National Institute of Child Health and Human Development Study of Early Child Care and Youth Development (NICHD SECCYD). The general goal is to advance understanding of child development, education, and inequality by treating schools as one piece of a web of overlapping contexts contributing to the learning of children from different backgrounds and assessing the give and take between selection and socialization in this web.

The Transition Into Elementary School and Socioeconomic Inequality

The transition into elementary school is a vehicle for socioeconomic stratification in the United States (Winsler et al., 2008). At this point, the educational system magnifies existing disparities in cognitive and academic skills. Because perceived skills set class placements and performance expectations that then build on each other, even small early differences between child groups can expand into large end-of-school outcomes, such as dropout rates (Entwisle, Alexander, & Olson, 1997; Raver, Gershoff, & Aber, 2007). These findings are part of the rationale that early investments provide the greatest long-term payoff and support policy efforts targeting preschool and primary education (Fuller, 2007; Heckman, 2006; Takanishi & Bogard, 2007).

Ecological and life course perspectives elucidate the role of the transition into school in socioeconomic inequality by emphasizing how it is situated in the multifaceted ecology of early childhood (Alexander & Entwisle, 1988; Burchinal, Campbell, Bryant, Wasik, & Ramey, 1997; Crosnoe, 2006b; Raver et al., 2007). One example of this approach is contextual systems theory (Pianta & Kraft-Sayre, 2003), which contends that learning during the transition into elementary school is best understood as either supported or disrupted by transactions among environmental settings. Learning proceeds more smoothly when these settings provide positive resources in a parallel or linked manner, less so when they are disorganized and/or misaligned. Importantly, the theory also asserts that differences in such transactions help to explain why some children, such as those from families of low socioeconomic status (SES), are at risk for poor transitions.

For the most part, the ecological settings studied in relation to the transition into elementary school are families and schools (Christenson & Sheridan, 2001; Epstein et al., 2002; Patrikakou, Weisberg, Redding, & Walberg, 2005). Contextual systems theory emphasizes multiple forms of transaction between the two. One form is direct and interactional, with parents and school personnel engaging each other, for example, through a regular pattern of mutually respectful parent–teacher meetings or when parents volunteer at schools and schools offer services to parents. Another form is less coordinated, for example, parents and school personnel cultivating children's skills in similar, mutually reinforcing ways (e.g., both reading to child) whether they know that they are doing so or not. A child experiencing this type of transaction is growing up in a system of *consistently* enriching settings. Evidence from the Early Childhood Longitudinal Study indicates that both family–school transactions are associated with learning gains in early childhood but that the presence of environmental consistency (even when not coordinated) is more closely related to socioeconomic disparities in early learning (Crosnoe, 2006a). This study, therefore, focuses on consistency in learning environments around the transition into school and addresses two potential areas of expansion in this line of research: including child care and unpacking the role of family SES.

Expanding the Scope

More fully recognizing the insights of ecological perspectives requires the consideration of more than just family–school connections. Given the aforementioned value of early educational investments (Heckman, 2006), early child care is a logical setting to examine. Today, most children transition into elementary school after spending time in nonparental child care, including formal center care. The calibration and assets provided by families, schools, and child care, then, create a web of possible support for children as they enter the educational system (Committee on Family and Work Policies, 2003). Just as research on family–school connections has yielded insight into the value of consistent learning supports across home and school (Epstein et al., 2002), both naturalistic and intervention studies have demonstrated the value to learning of positive transactions between home and child care/preschool (NICHD ECCRN, 2005; Wasik, Ramey, Bryant, & Sparling, 1990). What is lacking is information about coordination and/or consistency among all three pieces of the triangle.

Following contextual systems theory, the *accumulation* of cognitively stimulating settings should boost children's prospects for learning in elementary school. In other words, children who are stimulated at home, in child care, and then in elementary school would be expected to do better than children who are stimulated in only two of these settings, who would be expected to do better than children who are stimulated in only one of these settings. Beyond the accumulation of cognitively stimulating environmental settings, the *configuration* of exposures to cognitive stimulation matters. For example, children may not benefit equally from stimulation in two of three settings if the two settings in question differ between them (e.g., home and school for Child 1, child care and school for Child 2).

Given that parents are typically the most consistent agents of young children's learning and given the many ways that parents prepare children for school (Hart & Risley, 1995; Lareau, 2004), the absence of cognitive stimulation at home during early childhood should be especially problematic. Stimulation in *either* child care *or* school only, therefore, would be less beneficial than stimulation at home only. Furthermore, stimulation in *both* child care *and* school only would be less beneficial than stimulation in both home and child care only or in both home and school only. If stimulation at home is paired with stimulation in only one other setting, a case could be made that either school or child care will be more important. Of course, the school classroom is where formal instruction takes place, and, again, the demands of such instruction are what many parents are tailoring their in-home stimulation to eventually support (Entwisle et al., 1997). Yet, given the increasingly strong link between early child care and school readiness (NICHD ECCRN, 2005), children may be less able to take advantage of stimulating school settings if they do not enter school with a strong foundation for learning developed during years of consistent stimulation at home and in child care (Walker, Greenwood, Hart, & Carta, 1994).

The first aim of this study, therefore, is to identify configurations of cognitive stimulation across home, early child care, and school settings around the initial school transition. The second is to determine the degree to which these configurations are associated with children's learning trajectories during elementary school.

Elucidating the Role of Family SES

Given what is known about family SES and education (Entwisle et al., 1997; Walker et al., 1994), any assessment of consistent environmental stimulation during this transition would be incomplete without assessing how family SES fits into the puzzle. Theoretical perspectives on socioeconomic disparities in all stages of education have a basic underlying argument (Alexander & Entwisle, 1988; Lareau, 2004). In short, because socioeconomic disparities in access to social resources for learning drive socioeconomic differences in learning, equal access to such resources across socioeconomic strata should narrow these differences. This argument connects to the aims of this study in two ways.

First, one arrangement of family SES, social resources for learning, and child learning is that social resources mediate the link between family SES and child learning. An alternate explanation, however, is that the association between resources and learning is spurious, with both produced by family SES. In this case, social resources for learning are merely proxies for families' socioeconomic circumstances. Of course, both arrangements could be at work. The degree to which both, either, or neither apply here, then, needs to be determined.

Second, the link between consistent environmental stimulation and learning may also be moderated by SES. Again, the working hypothesis has long been that reducing group differences in access to social resources will reduce disparities in child outcomes. Yet, universal interventions often reveal that equal access to resources can lead to greater disparities in outcomes when non-at-risk children are better able to capitalize on having that resource than at-risk children (Ceci & Papierno, 2005). Following this pattern of cumulative advantage, high-SES children might benefit more from having the same kind of environmental stimulation as low-SES children. If so, socioeconomic disparities in learning would increase even if the absolute levels of learning of low-SES children were raised. Another scenario is that access to any new resource will make more of a difference to the children for whom it is nonredundant. Thus, the impact of a resource has added value for at-risk children. Following this buffering pattern, making sure that high- and low-SES children have the same kind of environmental stimulation should reduce socioeconomic disparities (Winsler et al., 2008). In other words, two moderation patterns are plausible.

The final aim of this study, therefore, involves three related questions. Are relations between family SES and child learning mediated by consistent environmental stimulation? Does consistent stimulation proxy family SES? Does SES condition effects of consistent stimulation?

Method

Data Source

Data collection for the SECCYD (http://secc.rti.org) began in 1991 in 10 locations: Little Rock, AR; Irvine, CA; Lawrence, KS; Boston, MA; Philadelphia, PA; Pittsburgh, PA; Charlottesville, VA; Morganton, NC; Seattle, WA; Madison, WI. During selected sampling periods, study personnel visited new mothers in the hospital. To be eligible, the mother had to be 18 years of age or older, healthy, and conversant in English, and the infant had to be a singleton. A month later, 1,364 families were enrolled in the study. The collection of data from parents, children, and other adults in home, laboratory playroom, child care, and school visits proceeded in several stages from birth and is ongoing. Because of our focus on the transition into school, we drew on the SECCYD data through the children's fifth-grade year. Some families dropped out or were lost by this time point (*n* = 287), and others missed some data collections. Because deleting cases with item- or wave-specific missingness can bias the sample, full information estimation methods were employed to allow for the retention of all observed data.

Measures

Child learning. Two Woodcock–Johnson Psycho-Educational Battery–Revised subtests (WJ–R; Woodcock & Johnson, 1989) were assessed in the laboratory at 54 months and Grades 1, 3, and 5. Letter–Word Identification gauged reading skills, and Applied Problems gauged math skills. Raw scores on the subtests were converted to *W* scores, transformations of the Rasch ability scale that centered the raw score at 500 in order to ease comparisons across the multiple WJ–R administrations. Table 1 includes mean scores on these subtests at each time point (as well as descriptive statistics for the other study variables). Scores increased across grades.

Family SES. Three SES measures tapped characteristics of the families themselves, reported by parents when children were 1 month old. Parent education was a binary marker (1 = *at least one parent had a college degree*, 0 = *other*). Income-to-needs was a continuous variable representing total income divided by the federal poverty threshold for a family of that size (U.S. Census Bureau, 2008). Marital status was a binary marker (1 = *parents married*, 0 = *other*). A fourth SES measure tapped collective characteristics of families in children's

Table 1
Descriptive Statistics for Study Variables (N = *1,364*)

Child, family, and setting characteristics	*M*	*SD*
Child test scores		
Reading score (54 months)	424.72	19.72
Reading score (Grade 1)	470.05	15.54
Reading score (Grade 3)	493.50	12.76
Reading score (Grade 5)	510.54	13.04
Math score (54 months)	369.36	21.41
Math score (Grade 1)	452.59	23.99
Math score (Grade 3)	494.60	15.75
Math score (Grade 5)	507.63	14.12
Family socioeconomic circumstances		
Parent education (college graduate)	0.44	0.50
Family income-to-needs	3.63	3.17
Parent marital status (married)	0.77	0.42
Neighborhood affluence	−0.04	0.89
Environmental stimulation		
HOME stimulation (54 months)	17.84	2.47
Child-care stimulation (54 months)	2.21	0.80
School classroom stimulation (Grade 1)	15.67	4.16
Other child/family factors		
Gender (male)	0.52	0.50
Race (non-White)	0.20	0.40
School readiness (36 months)	9.02	2.89
Maternal employment	0.67	0.47
Maternal employment (unemployed)	0.31	0.46
Maternal employment (on leave)	0.02	0.15

Note. For binary measures (parent education, marital status, gender, race, maternal employment), means represent the percentage of respondents in the one category. HOME = Home Observation of Measurement of the Environment.

neighborhoods. Neighborhood affluence was the mean of two standardized census block variables: (a) the percentage of professionally educated persons over the age of 25 in the block group and (b) the median household income in the block group ($\alpha = .89$). The descriptive statistics in Table 1 reveal that the sample exceeds national averages on many socioeconomic indicators. Ancillary analyses revealed a great degree of stability in these measures over time.

Environmental stimulation. Our typology was based on three *observational* indicators of environmental cognitive stimulation. First, cognitive stimulation in the family at 54 months was measured with the Home Observation of Measurement of the Environment (HOME; Bradley & Caldwell, 1979). The HOME enrichment scale summed the learning materials (sum of 11 binary items), academic stimulation (5 items), and variety (9 items) subscales ($\alpha = .88$). Second, cognitive stimulation in child care at 54 months was based on the Observational Rating of the Care Environment (ORCE; NICHD Early Child Care Research Network., 2005). The ORCE stimulation of cognitive development subscale was defined by ratings of high-quality stimulation provided in lessons, explanations, and activities (1 = *not at all characteristic*, 4 = *highly characteristic*). Third, cognitive stimulation in the first-grade classroom was based on the Classroom Observation System (COS; Pianta, Belsky, Houts, Morrison, & the NICHD ECCRN, 2007). The instructional quality composite ($\alpha = .70$) summed four ratings (1 = *uncharacteristic*, 7 = *extremely characteristic*): literacy, evaluative feedback, instructional conversation, and child responsibility.

To capture accumulation and configuration of environmental stimulation, these three scales needed to be categorized into levels and then combined in theoretically meaningful ways. After extensive preliminary analyses, the most stimulating contexts were identified as those scoring above the median on the HOME, ORCE, and COS. The final typology then represented the different combinations of 0–3 high-stimulation environments. Cross-classifying these binary measures created dummy variables tapping six categories of environmental consistency: (a) triple stimulation (32%), which included children above the median on all three; (b) family–child care stimulation (15%), which included children above the median on the HOME and ORCE only; (c) family–school stimulation (11%), which included children above the median on the HOME and COS only; (d) child care–school stimulation (9%), which included children above the median on the ORCE and COS only; (e) single stimulation (26%), which included children above the median on either the HOME, ORCE, or COS but not the other two; and (f) low stimulation (7%), which included children below the median on all three.

Other child and family factors. Several control variables were created to account for major observable confounds, including binary markers of gender (1 = *male*) and race (1 = *non-White*), maternal employment dummy variables (employed for pay part- or full-time, unemployed, on leave from work), and a continuous measure of school readiness at 36 months. This last measure was based on the Bracken (1984) Basic Concept Scale, a composite of five cognitive assessments of children: identifying and recognizing colors, letters, numbers, comparisons, and shapes (range = 0–61, $\alpha = .93$).

Procedure

As a first step, growth curves for reading and math test scores were estimated with free-loading latent curve models (Meredith & Tisak, 1990), which did not impose any specific trend over time but instead allowed the data to dictate it. The four WJ–R scores (54 months; Grades 1, 3, 5) served as indicators of latent variables representing the intercept and slope of the trajectory of reading or math achievement. The model was identified with constraints on the Grade 1 and 5 loadings. Specifically, the Grade 1 loading was constrained to 0 to allow the intercept to be interpreted as the mean level of achievement in Grade 1. Note that the intercept was situated at Grade 1 rather than 54 months so that the parallel learning categories (some of which contained the classroom item measured in Grade 1) could be meaningful predictors of the intercept. Next, the Grade 5 loading was constrained to 1 to allow the slope to be interpreted as the change in achievement between Grades 1 and 5. All other loadings were freely estimated. These loadings can be interpreted as a proportion of change in reading or math test scores based on the total change between Grades 1 and 5. For example, a Grade 3 loading of .60 would indicate that 60% of the total change in reading or math test scores between Grades 1 and 5 was achieved by Grade 3.

Next, environmental stimulation categories, family SES, and the child and family controls were entered into the baseline model as predictors of the intercept and slope of reading and math achievement. Comparing how results for socioeconomic circumstances and for environmental stimulation changed with the addition of the other two sets of factors to the model assessed the extent to which the mediational or spurious arrangements held in this sample. As a final step, interactions between the family SES and environmental stimulation variables were added to the growth curve model as predictors of the intercept and slope to explore moderation.

In all models, parameter estimates were obtained with full information maximum likelihood (FIML) in Mplus (Muthén & Muthén, 2004). Models were evaluated with several fit indices: χ^2, root mean square error of approximation (RMSEA), the Tucker–Lewis index (TLI), and the comparative fit index (CFI).

Results

As a starting point, Table 2 presents mean family socioeconomic characteristics for children in the six environmental stimulation categories. As expected, children in the triple stimulation category typically had the most advantaged family backgrounds, and those in the low stimulation category typically had the least advantaged family backgrounds. Of the four other environmental stimulation categories, family–child care and family–school stimulation had higher mean levels on the family socioeconomic factors than child care–school and single stimulation. Generally speaking, then, family socioeconomic circumstances and exposure to stimulating environments increased in tandem, but, perhaps not surprisingly, the link between family socioeconomic circumstances and stimulation at home appeared especially strong.

Of course, attempts to understand relations between SES factors and environmental stimulation categories are sensitive to the potential for sparse coverage in various SES × Environmental category cells. Cross-classifying the environmental stimulation typology with a set of categories for each SES indicator (parent married to child's other parent vs. not; parent college educated vs. not; income-to-needs of 0–2, 1–3, 3+; low, medium, and high neighborhood affluence defined by standard deviation cuts) revealed that some cells contained small numbers of children (e.g., 10 or fewer). The smallest cells were for children in the low stimulation category who had college-educated parents, middle-income parents, neighborhoods of low affluence, and neighborhoods of high affluence. The case coverage in these particular cells warrants caution when interpreting null or significant results.

Differences in Reading and Math Trajectories

Before assessing differences in test scores by environmental stimulation categories and family socioeconomic circumstances, we first had to determine the functional form of the test score trajectories across elementary school. Table 3 presents the results from unconditional free-loading growth curve models for reading and math.

For these unconditional models, fit was somewhat marginal for reading and moderate for math, although fit statistics improved in later modeling iterations. As explained earlier, the mean intercept for both subjects represents the average test score in Grade 1 (even though the 54-month score was also included), and the mean slope represents the average total change in test scores between Grades 1 and 5. Thus, children scored about 450 points, on average, on the Grade 1 WJ–R test for reading and then added an additional 55 points by the end of Grade 5. The basic growth curve was similar for math, although children started off, on average, a

Table 2
Family Socioeconomic Circumstances, by Environmental Stimulation Categories

	M (*SD*)					
Family characteristics	Triple stimulation (*n* = 436)	Family–child-care stimulation (*n* = 205)	Family–school stimulation (*n* = 150)	Child care–school stimulation (*n* = 123)	Single stimulation (*n* = 355)	Low stimulation (*n* = 95)
Parent education (college graduate)	0.67 (0.47)	0.58 (0.50)	0.59 (0.50)	0.36 (0.48)	0.39 (0.49)	0.31 (0.47)
Family income-to-needs	4.85 (4.66)	4.27 (2.72)	4.80 (5.97)	3.44 (2.95)	3.43 (2.60)	2.66 (2.56)
Parent marital status (married)	0.89 (0.31)	0.84 (0.37)	0.85 (3.63)	0.76 (0.43)	0.70 (0.46)	0.40 (0.50)
Neighborhood affluence	0.29 (0.94)	0.06 (0.89)	0.11 (0.89)	−0.03 (0.81)	−0.13 (0.83)	−0.33 (0.76)

Note. For binary measures (parent education, marital status), means represent the percentage of respondents in the one category.

Table 3
Results from Unconditional Growth Curve Models of Reading and Math Scores (N = 1,364)

	Reading		Math	
Model parameters	Intercept	Slope	Intercept	Slope
Mean (*SE*)	450.34*** (1.16)	55.00*** (0.72)	468.42*** (0.96)	41.83*** (0.59)
Variance (*SE*)	215.52*** (10.29)	31.41*** (4.39)	157.41*** (8.01)	16.23*** (3.33)

Note. Chi-square (19 *df*) = 206.84*** (reading), 66.75*** (math); TLI = 0.93 (reading), 0.98 (math); CFI = 0.96 (reading), 0.96 (math); RSMEA = 0.08 (reading), 0.05 (math); 90% CI = 0.081–0.104 (reading), 0.035–0.059 (math).
$^{***}p < .001$.

little higher on this WJ–R test and gained slightly fewer points over time. For both reading and math, most of the upward change in the slope occurred between Grades 1 and 3. Worth noting is that the significant variance estimates indicate that children in the sample varied significantly around the mean intercept and slope for both reading and math. Such significant variation can then potentially be explained by independent variables, including the environmental stimulation dummy variables and family socioeconomic factors. Below, we present the results for reading test scores in detail and then later summarize the findings for math test scores.

Consistent Environmental Stimulation, Family SES, and Reading

The first set of models in Table 4 (see Model 1) included the environmental stimulation dummy variables as the only predictors of the intercepts and slopes of the growth curve. The stimulation coefficients are interpretable as test point differences relative to the reference, child care–school stimulation, which was chosen because it provided the contrast with the largest number of observable significant differences among categories. Of note is that post hoc between-group comparisons were performed (based on the corresponding model-implied parameter estimates and standard errors) to catalog *all* pairwise differences among categories—not just those in relation to the reference category. Thus, the reference is important for interpreting coefficients in the tables but not important for understanding the full set of patterns analyzed.

For the reading intercept, the model coefficients and post hoc comparisons together indicated a basic rank ordering (from highest to lowest) in average Grade 1 reading scores: (a) triple stimulation, family-child-care stimulation; (b) family–school stimulation; (c) child care–school stimulation, single stimulation; (d) single stimulation, low stimulation. In this rank ordering, an environmental stimulation category that is in two separate ranks is a "bridge" between ranks. For example, the coefficient for single stimulation was not significantly different from the child care–school stimulation coefficient or the low stimulation coefficient, but the child care–school stimulation and low stimulation coefficients did differ from each other. In other words, single stimulation bridged two different ranks.

For the slope, the positive low stimulation coefficient in the table indicated that children in this environmental stimulation category had more pronounced reading gains after Grade 1 than children in the child care–school stimulation category. Post hoc comparisons revealed that they also had larger reading gains than children in *all* other environmental stimulation categories. Basically, children in the low stimulation category (i.e., those with the least exposure to stimulation overall) appeared to benefit the most from the start of formal schooling. They had the greatest uptick in reading skills after school had begun. This pattern is interesting given that the classrooms they entered once school began were not high in cognitive stimulation.

Although our environmental categorization scheme was aligned with our conceptual emphasis on accumulation and configuration, we recognized that categorizing and combining continuous variables could raise concerns about loss of scale variation. Consequently, we estimated this model (and subsequent models) with different categorization schemes (e.g., with standard deviation cut points), finding no difference in results for the largest categories. Moreover, the model was estimated with two- and three-way interactions among the continuous versions of the HOME, ORCE, and COS scales. Doing so revealed weaker associations between links among the three setting factors and

Table 4

Results From Growth Curve Models of Reading Scores by Environmental Stimulation Categories and Family SES (N = 1,364)

Predictors	*B* (*SE*) Model 1		*B* (*SE*) Model 2		*B* (*SE*) Model 3		*B* (*SE*) Model 4	
	Intercept	Slope	Intercept	Slope	Intercept	Slope	Intercept	Slope
Environmental stimulation								
Triple stimulation	8.93*** (1.56)	−1.08 (0.77)			5.78*** (1.45)	−0.70 (0.78)	4.37*** (1.22)	−0.33 (0.75)
Family–child care stimulation	6.14* (1.79)	−0.52 (0.88)			3.87* (1.63)	−0.31 (0.88)	2.51† (1.37)	−0.29 (0.85)
Family–school stimulation	3.03† (1.69)	−0.97 (0.83)			1.69 (1.54)	−0.82 (0.82)	0.87 (1.29)	−0.71 (.79)
Child care–school stimulation	—	—			—	—	—	—
Single stimulation	−0.84 (1.57)	0.30 (0.77)			−0.23 (1.42)	0.23 (0.76)	−0.40 (1.19)	0.21 (0.73)
Low stimulation	−5.10* (2.20)	3.73*** (1.08)			−2.82 (2.04)	3.51*** (1.09)	−0.5 (1.72)	3.10*** (1.05)
Family socioeconomic circumstances								
Parent education (college graduate)			7.57*** (1.03)	−1.25* (0.56)	6.82*** (1.02)	−1.10* (0.56)	3.22*** (0.87)	−0.13 (0.55)
Family income-to-needs			0.45** (0.16)	−0.09 (0.08)	0.37* (0.15)	−0.07 (0.08)	0.10 (0.13)	−0.01 (0.08)
Parent marital status (married)			4.08*** (1.08)	0.50 (0.60)	2.75* (1.10)	1.07 (0.61)	1.33 (0.96)	0.73 (0.60)
Neighborhood affluence			2.82*** (0.59)	−0.75* (0.32)	2.38*** (0.58)	−0.61† (0.32)	0.80 (0.49)	−0.33 (0.31)
Other child and family factors								
Gender (male)							0.39 (0.69)	0.45 (0.43)
Race (non-White)							0.00 (1.01)	−2.71*** (0.64)
School readiness (36 months)							2.91*** (0.14)	−0.80*** (0.09)
Maternal employ (unemployed)							−4.32† (2.22)	−0.73 (1.38)
Maternal employ (on leave)							−0.48 (0.75)	0.35 (0.48)

Note. Chi-square/*df* = 212.47/29*** (M1), 215.00/27*** (M2), 219.82/37*** (M3), 224.24/47*** (M4); TLI = 0.94 (M1), 0.98 (M2), 0.94 (M3), 0.95 (M4); CFI = 0.87 (M1), 0.96 (M2), 0.88 (M3), 0.90 (M4); RSMEA = 0.07 (M1), 0.04 (M2), 0.05 (M3), 0.05 (M4); 90% CI = 0.063–0.081 (M1), 0.029–0.049 (M2), 0.053–0.068 (M3), 0.047–0.060 (M4).

$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

the outcomes, which was likely attributable to the nonlinear nature of the scales in combination with each other.

The second set of models in Table 4 (see Model 2) included the family socioeconomic factors as the only predictors of the intercepts and slopes of each growth curve (i.e., no environmental stimulation categories or family/child controls). All four significantly predicted the reading intercept. The difference between children with and without a college-educated parent equaled about 7.5 points on the reading test in Grade 1. The spread was .5 point for each increase in the income-to-needs and 4 points between children with and without married parents. Standardizing all four coefficients revealed that the effect size for neighborhood affluence was between the effect sizes for parent education (the biggest) and family income-to-needs (the smallest). As for the reading slope, two family socioeconomic factors had significant, negative coefficients. Children with college-educated parents and children who lived in more affluent neighborhoods had slightly less pronounced upward rates of change in achievement after Grade 1. Again, the most disadvantaged appeared to benefit the most from formal schooling.

Assessing Mediation

Mediation entails differences in child learning by family SES channeled through corresponding SES differences in environmental stimulation. One way to assess mediation is to compare associations between socioeconomic circumstances and child test scores before and after taking environmental stimulation into account. If environmental stimulation mediates the associations between socioeconomic circumstances and child test scores, then these associations should be weaker after environmental stimulation is controlled. Assessing the degree to which they are weaker gauges the degree of mediation.

The baseline associations between the four family socioeconomic circumstances and the reading achievement intercept and slope appear in Model 2 of Table 4. Model 3 reveals what happened to these baseline associations once the full set of environmental stimulation dummy variables was added to the growth curve model without any other controls.

Beginning with the previously observed association between parent education and the reading intercept, the parent education coefficient in Model 2 was attenuated by 10% with the inclusion of the environmental stimulation dummy variables in Model 3. Because triple stimulation and family-child-care stimulation were the only significant coefficients among the environmental stimulation dummy variables in Model 3, they appeared to be driving this attenuation. The corresponding levels of attenuation between intercept Models 2 and 3 were 18% for family income, 33% for parent marital status, and 16% for neighborhood affluence. Thus, some of the advantage that children from more privileged backgrounds had over their peers in reading at the start of school appeared to be related to their greater likelihood of having stimulation at home matched with stimulation in child care and/or Grade 1 classrooms.

Turning to the slope, the previously observed negative associations of parent education and neighborhood affluence with the reading slope (Model 2) were slightly attenuated by the inclusion of the environmental stimulation dummy variables. Because low stimulation was the only stimulation coefficient to achieve statistical significance in Model 3, it appeared to be driving this attenuation. Thus, the shallower reading gains that children from more privileged backgrounds exhibited across the elementary school years were, in part, related to their tendency to have stimulation in any or all settings of early childhood; in other words, to *not* be in the low stimulation category. They started off ahead and then had less to gain over time.

Model 4 added the full set of child and family controls, of which early school readiness was the only one to significantly predict both intercept and slope. More school-ready children had higher initial reading test scores but slightly less pronounced test score gains as elementary school progressed. The inclusion of the Bracken measure attenuated the associations between the family socioeconomic circumstances and the reading intercept by between an additional one half and three fourths (comparing the intercept coefficients for the socioeconomic factors in Model 3 to the corresponding coefficients in Model 4). Along with race, the inclusion of the school readiness scale also washed out the observed associations between parent education and neighborhood affluence on one hand and the reading slope on the other (comparing the slope coefficients for the socioeconomic factors in Model 3 and Model 4).

Assessing Spuriousness

Spuriousness occurs when test score differences across environmental stimulation categories are accounted for by the family socioeconomic circumstances that contribute to both environmental stimulation and test scores. One way to assess spuriousness is to compare associations between environmental stimulation categories and test scores before and after taking family socioeconomic circumstances into account. If these associations are spurious (confounded with family socioeconomic circumstances), then they should be weaker after family socioeconomic circumstances are controlled.

Again, comparing coefficients between models allows an assessment of spuriousness (see Table 4). The baseline associations between the environmental stimulation dummy variables and the reading intercepts and slopes have already been presented in Model 1. The coefficients in Model 3, which simultaneously regressed the intercept and slope of reading on the family socioeconomic circumstances and the environmental stimulation dummy variables, provide the appropriate comparison. For the reading intercept, the environmental stimulation coefficients were smaller when they were estimated simultaneously with the family SES factors (Model 3 intercept) than when estimated in isolation (Model 1 intercept). The coefficients for triple stimulation and family–child care stimulation (relative to child care–school as well as single and low stimulation) remained significant but were attenuated. The coefficients for family–school stimulation and low stimulation were reduced by almost half and became nonsignificant.

When the other child and family controls were added in Model 4, the intercept coefficients for triple and family–child care stimulation appeared somewhat robust. Of the environmental stimulation dummy variables, they were the only two to be associated with the reading intercept after both the family socioeconomic circumstances and the child and family controls were taken into account. The slope coefficient for low stimulation was also robust to controls for family socioeconomic circumstances and other child and family characteristics.

Assessing Moderation

A final aim of this study was to investigate whether family socioeconomic circumstances and environmental stimulation were related to each other in more complex, multiplicative ways. Specifically, we were interested in whether the links between environmental stimulation categories and test scores varied as a function of family socioeconomic circumstances.

To assess moderation, the family socioeconomic factors, the environmental stimulation dummy variables, and their interactions were entered into the unconditional growth curve model as predictors of the reading intercept and slope. Interactions of the environmental stimulation dummy variables with each family socioeconomic factor were examined separately. Significant interactions were graphed to determine the degree to which the associations between environmental stimulation and reading test scores varied by family SES. These models were estimated with and without the set of child and family control variables. The results of revealed that, of the four socioeconomic factors, only family income-to-needs significantly interacted with the environmental stimulation categories in any consistent way. Table 5 presents the focal results of the reading models with the family income interactions. Model 1 was the baseline (i.e., main effects and interactions only), and Model 2 added the other child and family controls.

According to the results for Model 1, all environmental stimulation categories except low stimulation (recall the small cell size for Income × Low stimulation) significantly interacted with family income-to-needs to predict the reading intercept when child care-school stimulation was the reference. All of these interactions were negative in sign. To interpret these interactions, we calculated predicted intercept values for children in each environmental stimulation category (including the omitted category of child care–school stimulation) who came from low-income families (income-to-needs = 1) or high-income families (income-to-needs = 5), with all other variables held to their sample means (see Figure 1).

For high-income children, only triple stimulation was associated with a significantly higher reading intercept than the child care–school category reference. The difference between the two was 3 points, equaling 19% of a standard deviation in Grade 1 test scores (the time point that served as the intercept in the growth curve). The difference between triple stimulation and the other two environmental stimulation categories (family–school, family–child care) was 5 points (32% *SD*), although these two other categories did not significantly differ from the child care–school category. More differences across environmental categories were detected, however, for low-income children. For example, the predicted reading intercept for children with child care–school stimulation (444), the category without high family stimulation, was significantly lower than for children with family–school stimulation (3 points, 19% standard deviation) and family–child care stimulation (4, 26%). In turn, the predicted

Table 5

Selected Results From Reading Growth Curve Models With Family × Environmental Stimulation Category Interaction Terms to Assess Moderation (N = 1,364)

	B (*SE*) for Model 1		*B* (*SE*) for Model 2	
Predictors	Intercept	Slope	Intercept	Slope
Environmental stimulation				
Triple stimulation	8.85*** (2.11)	−0.63 (1.14)	6.57*** (1.78)	0.09 (1.10)
Family–child care stimulation	5.01* (2.28)	0.00 (1.24)	4.07* (1.92)	−0.20 (1.19)
Family–school stimulation	3.37 (2.32)	−0.30 (1.26)	3.41 (1.97)	−0.45 (1.21)
Child care–school stimulation	—	—	—	—
Single stimulation	0.10 (1.88)	1.11 (1.02)	−0.66 (1.58)	1.30 (0.97)
Low stimulation	−0.56 (2.70)	3.17* (1.46)	1.23 (2.29)	2.72 (1.41)
Family socioeconomic circumstances				
Family income-to-needs	1.84*** (0.46)	−0.15 (0.25)	0.80* (0.39)	0.09 (0.24)
Interaction terms				
Income × Triple Stimulation	−1.02* (0.50)	−0.12 (0.27)	−0.20 (0.43)	−0.31 (0.26)
Income × Family–Care Stimulation	−1.40* (0.65)	0.10 (0.35)	−0.75 (0.55)	−0.02 (0.34)
Income × Family–School Stimulation	−0.91* (0.38)	0.18 (0.21)	−0.73* (0.32)	0.13 (0.20)
Income × Care–School Stimulation	—	—	—	—
Income × Single Stimulation	−1.34* (0.57)	0.62* (0.31)	−0.76 (0.48)	0.48 (0.29)
Income × Low Stimulation	−0.14 (0.92)	−0.49 (0.49)	−0.08 (0.77)	−0.66 (0.47)

Note. Model 1 included the other SES indicators. Model 2 added the child and family factors. Chi-square/*df* = 236.78/57*** (M1), 224.24/67*** (M2); TLI = 0.94 (M1), 0.95 (M2); CFI = 0.88 (M1), 0.90 (M2); RSMEA = 0.05 (M1), 0.05 (M2); 90% CI = 0.042–0.055 (M1), 0.047–0.060 (M2).
**p* < .05. ****p* < .001.

intercept for children in these latter two categories, which tapped double stimulation when at least one stimulating environment was the home, was significantly lower than for children in the triple stimulation category (4 and 5 points, respectively, at 26% and 32% *SD*). Indeed, the biggest within-income difference in Figure 1 was between low-income children with child care–school stimulation and children with triple stimulation (8 points, 51% *SD*). Thus, among low-income children, each environmental stimulation category involving stimulation at home was associated with higher reading intercepts than child care–school stimulation.

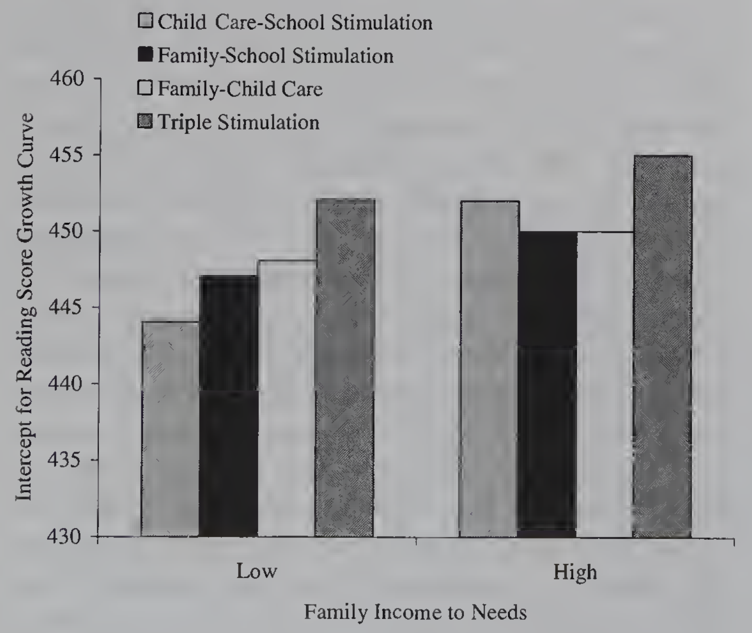


Figure 1. Reading scores at intercept, by family income and environmental stimulation category.

When the full set of child and family controls was included (see Model 2 in Table 5), this basic pattern of moderation held only for the family–school environmental stimulation category. In other words, the observed "benefits" of having triple stimulation or family–child care stimulation were equal for low-income and high-income children of similar levels of early school readiness, but low-income children appeared to benefit more from linking stimulation at home and school than high-income but similarly school-ready peers.

Summary of Math Findings

In general, the results from the math models were similar to those for reading (Table 6). In terms of the associations between environmental stimulation and test scores, one major distinction (see Table 4) emerged. It involved the math intercept.

Table 6

Main Results from Growth Curve Models of Math Scores (N = *1,364*)

	B (SE) Model 1		*B (SE)* Model 2		*B (SE)* Model 3		*B (SE)* Model 4	
Predictors	Intercept	Slope	Intercept	Slope	Intercept	Slope	Intercept	Slope
Environmental stimulation								
Triple stimulation	5.40*** (1.41)	−1.92* (0.77)			2.42 (1.29)	−1.45 (0.77)	1.24 (1.13)	−.99 (0.75)
Family–child care stimulation	6.16*** (1.60)	−0.71 (0.87)			3.93** (1.45)	−0.39 (0.87)	2.49* (1.26)	0.03 (0.84)
Family–school stimulation	3.20* (1.52)	−0.91 (0.82)			1.86 (1.37)	−0.71 (0.82)	1.14 (1.19)	−0.48 (0.78)
Child care–school stimulation	—	—			—	—	—	—
Single stimulation	−0.72 (1.41)	−0.13 (0.76)			−0.13 (1.26)	−0.21 (0.76)	−0.21 (1.09)	−0.29 (0.73)
Low stimulation	−6.30*** (1.99)	2.15* (1.08)			−3.76* (1.82)	1.76 (1.09)	−2.08 (1.58)	1.15 (1.05)
Family socioeconomic circumstances								
Parent education (college graduate)			5.91*** (0.89)	−0.40 (0.56)	5.50*** (0.89)	−0.24 (0.56)	2.71*** (0.79)	0.68 (0.55)
Family income-to-needs			0.50*** (0.14)	−0.16 (0.08)	0.45*** (0.14)	−0.13 (0.08)	0.20 (0.12)	0.06 (0.08)
Parent marital status (married)			4.86*** (0.94)	−0.40 (0.60)	3.81*** (0.97)	−0.01 (0.61)	2.19* (0.87)	0.67 (0.61)
Neighborhood affluence			2.51*** (0.50)	−0.87 (0.32)	2.26*** (0.50)	−0.75* (0.32)	0.99* (0.44)	−0.29 (0.31)
Other child and family factors								
Gender (male)							1.15 (0.63)	0.84 (0.45)
Race (non-White)							−2.04* (0.91)	1.24 (0.65)
School readiness (36 months)							2.22*** (0.13)	−0.72*** (0.09)
Maternal employ (unemployed)							−4.39* (2.02)	−1.66 (1.39)
Maternal employ (on leave)							−1.30 (0.68)	−0.21 (0.48)

Note. Chi-square/*df* = 79.76/29** (M1), 86.96/27*** (M2), 96.99/37*** (M3), 158.36/47*** (M4); TLI = 0.98 (M1), 0.98 (M2), 0.98 (M3), 0.96 (M4); CFI = 0.96 (M1), 0.95 (M2), 0.95 (M3), 0.92 (M4); RSMEA = 0.04 (M1), 0.04 (M2), 0.03 (M3), 0.04 (M4); 90% CI = 0.029–0.049 (M1), 0.031–0.050 (M2), 0.026–0.043 (M3), 0.035–0.049 (M4).

**p* < .05. ***p* < .01. ****p* < .001.

Children in the triple and family–child care stimulation categories had significantly higher scores in Grade 1 than children in all other categories. Moreover, in addition to the same low stimulation association seen in the reading models, triple stimulation also predicted the math slope. Children in this category had less pronounced gains in math after Grade 1 relative to those in the child care–school stimulation category and indeed relative to all other categories (according to post hoc comparisons).

Overall, the results of models examining the associations between family socioeconomic advantages and children's math trajectories were quite similar to the reading results in Table 4, with the one exception being that the association between parent education and the math slope was nonsignificant. Looking at mediation, the baseline associations between the family socioeconomic circumstances and the math intercept and slope were partially accounted for by the inclusion of the environmental stimulation dummy variables (family–child care and low stimulation for the intercept, low stimulation for the slope) and substantially mediated by the inclusion of the school readiness scale, as was the case for reading.

The only difference in the results of the spuriousness analyses for reading and math was that family–child care stimulation was the one environmental stimulation dummy variable to be associated with the math intercept before *and* after controlling for family socioeconomic circumstances and other child and family characteristics. Recall that, for reading, both the triple stimulation and family–child care stimulation demonstrated this pattern (see Table 4).

Finally, the results of math models testing for moderation (not presented but available upon request) were similar to the reading results presented in Table 5, except that family income also interacted significantly with single stimulation to predict the slope of the growth curve. Basically, the apparent "benefit" of stimulation in two or three early settings relative to only one was more pronounced for low-income children than high-income children, both when school readiness was and was not taken into account.

Discussion

At first glance, the findings of this study seem unsurprising. High-SES children had more exposure to stimulating environments that promote learning. On closer inspection, however, the findings highlight important qualifiers to established patterns in the literature. First, they address tradeoffs between additive (or cumulative) and configuration (or typology) approaches to environmental effects on children's learning. Second, they qualify for whom different types of resources matter. Both the number and configuration of early sources of cognitive stimulation were critical for children's early reading and math trajectories. Children who experienced cognitive stimulation in multiple settings of early childhood had higher rates of learning than their peers early in school, but only when one of these settings was the home. As elementary school unfolded, all children typically had similar rates of progress from their unequal starting points. Although children with the least cognitive stimulation across settings had slightly higher increases in learning from Grades 1 to 5 compared to their peers with more consistent stimulation, this advantage was small relative to the large gap that existed between them when school began. Thus, early learning disparities across environmental stimulation categories generally persisted over time.

Overall, linking stimulation in the family to stimulation in any other early childhood contexts—in child care and/or school for reading, in child care for math—seemed beneficial for children's learning. This means that the family is the first, but not the only, ingredient in early learning. Parents, after all, spend the most (and most consistent) time with their children across early childhood and play a leading role in preparing children for school (Hart & Risley, 1995; Lareau, 2004). Of course, this pattern reflects the consideration of *single* dimensions of family, child care, and school settings. At the same time, we acknowledge that the home environment and the selection of more formal settings for children may be capturing genetic relatedness between parents and their children. Still, these findings are noteworthy for several reasons.

In general, the results of this study suggest that targeting single settings as a means of improving children's learning might not be sufficiently powerful for establishing or deflecting trajectories of early achievement. Specifically, policies that focus solely on one setting, such as parenting programs, preschool interventions, or school standards movements, are unlikely to be as efficacious as those targeting consistency across multiple settings. This conclusion supports the tenets of contextual systems theory contending that multisetting approaches will add value to policy efforts, especially those aimed at at-risk populations. It also suggests that efforts to facilitate children's transitions between home and school should be deliberately expanded to include early child care and preschool. Doing so seems especially crucial for math achievement, for which cognitively stimulating child care more than stimulating entry-level school classrooms was what was needed to match with stimulating home environments to promote early learning.

A major component of this study concerned the multifaceted linkages among socioeconomic stratification and consistency of environmental stimulation. Findings revealed that children from more advantaged families started school with more developed skills than their more disadvantaged peers and that these disparities persisted through elementary school. Beyond replicating this well-documented phenomenon (Magnuson, Meyers, Ruhm, & Waldfogel, 2004), we also sought to elucidate whether consistent environmental stimulation explained the link between family SES and child learning (mediation), family SES accounted for the association between consistent stimulation and child learning (spuriousness), or some combination of both processes was occurring. The motivation for comparing these theoretical pathways was that policy efforts that seek to reduce academic disparities typically attempt to remedy the mechanisms through which socioeconomic advantages are thought to operate in the educational system rather than directly addressing family SES itself. Along these lines, having stimulation at home along with stimulation in one or two other early settings partially explained why higher SES children had higher overall skill levels. These findings generally concur with past research exploring the contextual resources through which family SES is associated with child learning (Brooks-Gunn & Duncan, 1997; Entwisle et al., 1997). Yet, this study extends that research by incorporating the broader constellation of settings through which SES has its reach rather than focusing on single settings or on multiple settings in an isolated way.

Importantly, evidence of spurious associations between configurations of stimulating settings and children's trajectories also emerged. That is, the consistency of environmental stimulation that children experienced around the transition into school was a proxy for socioeconomic circumstances more powerfully related to their learning. Yet, several exceptions existed. Triple stimulation (for reading) and family–child care stimulation (for reading and math) had potentially additive effects. Perhaps more importantly, moderation analyses revealed that the tendency for consistent environmental stimulation to simply be a proxy for family SES appeared to be less strong for children from low-income families. They appeared to gain more than high-income peers through school when early stimulation in extrafamilial settings was consistent with early stimulation at home. These findings are comparable to intervention efforts documenting higher achievement gains when at-risk children experienced more concentrated stimulation across contexts rather than in a single setting (Burchinal et al., 1997; Wasik et al., 1990). That low-income children benefited the most from stimulation at home in combination with other early childhood settings (especially entry-level school classrooms) is also in line with the intervention literature pointing to differential efficacy as a function of risk status (Burchinal et al., 1997; Winsler et al., 2008). Moreover, it is promising news for large-scale policy efforts targeting this population, including Head Start, public Pre-K, and No Child Left Behind.

In addition to those we have already discussed, several limitations of this study need to be addressed. Methodologically, although we examined family SES as a potential source of omitted variable bias in the association between consistency of environmental stimulation and children's learning, other unobserved factors, including parental efficacy and genetic traits, may account for these links. More sophisticated strategies to address unobservable confound should be utilized in the future if this line of research is to effectively inform policy intervention (Winship & Morgan, 1999). On a more conceptual level, our interest in connections among different early childhood settings focused on indirect, temporally spaced transactions rather than direct engagement between actors across settings. A comprehensive examination should include both direct and indirect forms and compare these two approaches to conceptualizing transactions among different settings of the early ecology. Finally, although the sample used in this study had many advantages for our study aims, it was not nationally representative. The generalizability of these findings, therefore, needs to be assessed in the future.

In sum, our results suggest that family socioeconomic advantages were a driving force behind children's exposure to cognitive stimulation in various settings and in their learning. Although this conclusion echoes criticisms of more direct parental involvement strategies in schools, it does not imply that policies should not have environmental foci, particularly in the case of low-income children.

Indeed, in the SECCYD, these children uniquely benefited from the accumulation of stimulating settings. Clearly, altering environmental stimulation through preschools (e.g., Head Start, public pre-K) and schools (e.g., No Child Left Behind), while challenging, is likely more feasible than policies to reduce family socioeconomic disparities, at least in the short term. Efforts to address academic differentials need to include families (not just educational contexts), should be coordinated, and should start before children enter school.

References

Alexander, K. L., & Entwisle, D. R. (1988). Achievement in the first two years of school: Patterns and processes. *Monographs of the Society for Research in Child Development, 53*, 1–157.

Bracken, B. A. (1984). *Bracken Basic Concept Scale.* San Antonio, TX: Psychological Corporation.

Bradley, R. H., & Caldwell, B. M. (1979). Home Observation for Measurement of the Environment: A revision of the preschool scale. *American Journal of Mental Deficiency, 84*, 235–244.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.

Brooks-Gunn, J., & Duncan, G. (1997). The effects of poverty on children. *Future of Children, 7*, 55–71.

Burchinal, M. R., Campbell, F. A., Bryant, D. M., Wasik, B. H., & Ramey, C. T. (1997). Early intervention and mediating processes in cognitive performance of children of low-income African American families. *Child Development, 68*, 935–954.

Ceci, S. J., & Papierno, P. B. (2005). The rhetoric and reality of gap closing: When the ''have nots'' gain but the ''haves'' gain even more. *American Psychologist, 60*, 149–160.

Christenson, S. L., & Sheridan, S. M. (2001). *School and families: Creating essential connections for learning*. New York: Guilford.

Committee on Family and Work Policies. (2003). *Working families and growing kids: Caring for children and adolescents*. Washington, DC: National Academy.

Crosnoe, R. (2006a, August). *Family-school partnerships, early learning, and socioeconomic inequality*. Paper presented at the annual meeting of the American Sociological Association, Montreal, Canada.

Crosnoe, R. (2006b). *Mexican roots, American schools: Helping immigrant children succeed.* Palo Alto, CA: Stanford University Press.

Duncan, G. J., Magnuson, K. A., & Ludwig, J. (2004). The endogeneity problem in developmental studies. *Research in Human Development, 1*, 59–80.

Elder, G. H., Jr. (1998). The life course as developmental theory. *Child Development, 69*, 1–12.

Entwisle, D. R., Alexander, K. L., & Olson, L. S. (1997). *Children, schools and inequality*, Boulder, CO: Westview.

Epstein, J. L. (2005). Attainable goals? The spirit and letter of the No Child Left Behind Act on parental involvement *Sociology of Education, 78*, 179–182.

Epstein, J., Sanders, M., Simon, B., Salinas, K., Jansorn, N., & Van Voorhis, F. (2002). *School, family, and community partnerships: Your handbook for action*, Thousand Oaks, CA: Corwin.

Fuller, B. (2007). *Standardized childhood: The political and cultural struggle over early education*, Palo Alto, CA: Stanford University Press.

Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday experiences of young American children.* Baltimore: Brookes.

Heckman, J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science, 312*, 1900–1902.

Lareau, A. (2004). *Unequal childhoods: Class, race, and family life.* Berkeley: University of California.

Magnuson, K. A., Meyers, M. K., Ruhm, C. J., & Waldfogel, J. (2004). Inequality in preschool education and school readiness. *American Educational Research Journal, 41*, 115–157.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika, 55*, 107–122.

Muthén, L. K., & Muthén, B. O. (2004). *Mplus user's guide.* Los Angeles: Author.

National Institute of Child Health and Human Development Early Child Care Research Network. (2005b). Early child care and children's development in the primary grades. *American Educational Research Journal, 43*, 537–570.

Patrikakou, E. N., Weisberg, R. P., Redding, S., & Walberg, H. J. (Eds.). (2005). *School-family partnerships for children's success*, New York: Teachers College Press.

Pianta, R. C., Belsky, J., Houts, R., Morrison, F., & the NICHD Early Child Care Research Network. (2007). Opportunities to learn in America's elementary classrooms. *Science, 315*, 1795–1796.

Pianta, R. C., & Kraft-Sayre, M. (2003). *Successful kindergarten transition: Your guide to connecting children, families, and schools.* Baltimore: Brookes.

Raver, C. C., Gershoff, E., & Aber, L. (2007). Testing equivalence of mediating models of income, parenting, and school readiness for White, Black, and Hispanic children in a national sample. *Child Development, 78*, 96–115.

Takanishi, R., & Bogard, K. (2007). Effective educational programs for young children: What we need to know. *Child Development Perspectives, 1*, 40–45.

U.S. Census Bureau. (2008). Poverty thresholds: 1991. Retrieved May 1, 2009, from http://www.census.gov/hhes/www/poverty/threshld/thresh91.html.

Walker, D., Greenwood, C., Hart, B., & Carta, J. (1994). Prediction of school outcomes based on early language production and socioeconomic factors. *Child Development, 65*, 606–621.

Wasik, B. H., Ramey, C. T., Bryant, D. M., & Sparling, J. J. (1990). A longitudinal study of two early intervention strategies: Project CARE. *Child Development*, *61*, 1682–1696.

Winship, C., & Morgan, S. L. (1999). The estimation of causal effects from observational data. *Annual Review of Sociology*, *25*, 696–706.

Winsler, A., Tran, H., Hartman, S. C., Madigan, A. L., Manfra, L., & Bleiker, C. (2008). School readiness gains made by ethnically diverse children in poverty attending center-based childcare and public school pre-kindergarten programs. *Early Childhood Research Quarterly*, *23*, 314–329.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock–Johnson Psycho-Educational Battery–Revised*. Itasca, IL: Riverside.

Child Development, May/June 2010, Volume 81, Number 3, Pages 988–1005

Parent Involvement and Children's Academic and Social Development in Elementary School

Nermeen E. El Nokali, Heather J. Bachman, and Elizabeth Votruba-Drzal
University of Pittsburgh

Data from the National Institute of Child Health and Human Development (NICHD) Study of Early Childcare and Youth Development (N = 1,364) were used to investigate children's trajectories of academic and social development across 1st, 3rd, and 5th grades. Hierarchical linear modeling was used to examine within- and between-child associations among maternal and teacher reports of parent involvement and children's standardized achievement scores, social skills, and problem behaviors. Findings suggest that within-child improvements in parent involvement predict declines in problem behaviors and improvements in social skills but do not predict changes in achievement. Between-child analyses demonstrated that children with highly involved parents had enhanced social functioning and fewer behavior problems. Similar patterns of findings emerged for teacher and parent reports of parent involvement. Implications for policy and practice are discussed.

A recent report by Herrold and O'Donnell (2008) from the National Center for Education Statistics found that over 90% of parents of elementary school children reported attending general school meetings, like those for the PTA or PTO, as well as participating in regularly scheduled parent teacher meetings throughout the school year. Roughly 80% of parents in the nationally representative sample reported attending school events and about 60% reported volunteering in the classroom. Parent involvement is often considered a pathway through which schools enhance the achievement of underperforming children (Berger, 1991). Consequently, parent involvement is encouraged by teachers, child-care providers, policy makers, parents, and researchers (Duch, 2005; Sheldon & Epstein, 2005). Moreover, parent involvement is a key component of early childhood education programs, such as Head Start. These programs encourage parent involvement by inviting parents to participate in activities at school and facilitating parent–teacher communication.

Over the years, a large literature has documented the importance of parent involvement for young children. The role of parent involvement in the later years of schooling has received less attention. Past research on parent involvement has also been more heavily focused on associations with student achievement, with less attention to social and emotional domains of children's development. This propensity may be attributed to the academic nature of many of the behaviors defined as parent involvement like helping with homework. Such activities should prompt more enrichment at home and attunement to a child's academic progress. However, teachers and parents may discuss children's behavior in the classroom as well, since behavior problems and social functioning may have immediate consequences for the classroom environment and teachers' instructional efforts. The aim of this study is to extend past research by examining within- and between-child associations among parent involvement and children's academic and socioemotional trajectories during elementary school.

This research was supported by a grant to Heather J. Bachman and Elizabeth Votruba-Drzal from the National Institute of Child Health and Human Development (NICHD; 1RO3HD056019-01A1). This study was conducted by the NICHD Early Child Care Research Network supported by NICHD through a cooperative agreement that calls for scientific collaboration between the grantees and the NICHD staff. An earlier version of this article was presented at the 2009 SRCD Conference.

We would like to thank our discussant, Eric Dearing, as well as the anonymous referees for their comments. The content is solely the responsibility of the authors and does not necessarily reflect the views of the NICHD. We would also like to thank Roli Mohan, Stephanie Davis, and Carolina Maldonado-Carreño for their assistance with data management. A special thanks is also extended to the children and families who participated in the NICHD Study of Early Child Care and Youth Development.

Correspondence concerning this article should be addressed to Nermeen E. El Nokali, Department of Psychology in Education, School of Education, University of Pittsburgh, 5930 Posvar Hall, Pittsburgh, PA 15260. Electronic mail may be sent to nee5@pitt.edu.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8103-0023

Parent Involvement in Early Childhood

Parent involvement typically involves parents' behaviors in home and school settings meant to support their children's educational progress. Measures of parent involvement commonly include the quality and frequency of communication with teachers as well as participation in school functions and activities (Dearing, Kreider, Simpkins, & Weiss, 2006; Dearing, McCartney, Weiss, Kreider, & Simpkins, 2004; Machen, Wilson, & Notar, 2004). Parent involvement also characterizes parents' values and attitudes regarding education and the aspirations they hold for their children (Catsambis, 2001; Englund, Luckner, Whaley, & Egeland, 2004). Although values and attitudes may not directly influence academic outcomes, they may enhance academic achievement indirectly by promoting children's motivation and persistence in challenging educational tasks.

Parent involvement bridges two key contexts in children's early development, namely, the home and school settings. Within an ecological framework (Bronfenbrenner & Ceci, 1994), the home and school contexts are characterized as autonomous microsystems and parent involvement is conceptualized as a mesosystem, which is made up of interactions between key microsystems. Although each setting can independently influence a child, together the home and school contexts interact to offer a unique influence. In this study parent involvement is conceptualized as a product of the interaction between the influences of school and home settings by providing continuity between the two environments. For example, if parents are aware of a teacher's instructional goals, they may provide resources and support for those learning aims at home. Similarly, in terms of social development, parent involvement may facilitate the development of consistent disciplinary approaches across home and school. Accumulating evidence suggests that these parenting practices are associated with higher academic success in the early grades, although links to socioemotional outcomes remain less clear.

Academic Achievement

Past research on parent involvement and children's academic skills is mixed (Fan & Chen, 2001). Some studies have found no significant association between parent involvement and academic achievement (Keith, Reimers, Fehrmann, Pottebaum, & Aubey, 1986; Okpala, Okpala, & Smith, 2001; Reynolds, 1992; White, Taylor, & Moss, 1992) and a few have even detected negative associations (Milne, Myers, Rosenthal, & Ginsburg, 1986; Sui-Chu & Willms, 1996). Yet, positive associations between parent involvement and academic achievement have been demonstrated repeatedly in the literature. A recent meta-analysis by Fan and Chen (2001) finds moderate associations between parent involvement and an array of learning-related or academic skills, such as achievement motivation, task persistence, and receptive vocabulary, during preschool and kindergarten. With a predominant research focus on parent involvement and achievement in either preschool and kindergarten or high school, the potentially supportive role of parent involvement during middle childhood remains understudied.

Past nonexperimental research on parent involvement commonly investigates contemporaneous associations between parent involvement and academic achievement. These studies typically examine within-grade associations of parent involvement and academic skills (Fantuzzo, McWayne, & Perry, 2004; Gonzalez-DeHass, Willems, & Holbein, 2005; McWayne, Hampton, Fantuzzo, Cohen, & Sekino, 2004). Other work incorporates contemporaneous research in the early grades with longitudinal follow-up data later in elementary school or beyond (Englund et al., 2004; Izzo, Weissberg, Kasprow, & Fendrich, 1999; Miedel & Reynolds, 1999). For example, Miedel and Reynolds (1999) detected positive associations between parent involvement in preschool and kindergarten and reading achievement in kindergarten and in eighth grade. Izzo et al. (1999) also found significant positive associations between average parent involvement in early elementary school and socioemotional development and achievement in later elementary school. Such studies reflect the common practice of considering parent involvement as a static predictor of concurrent achievement or educational outcomes in later school years. A notable exception is a study by Dearing et al. (2006) that employed longitudinal data on parent involvement and reading achievement to examine within- and between-family associations of parent involvement and literacy across elementary school. Findings suggested that differences in levels of parent involvement between families and changes in parent involvement within families were both predictive of children's literacy skills, and increasing parent involvement during elementary school improved literacy growth.

Socioemotional Development

Parent involvement is generally thought of as an avenue for promoting academic performance. However, parent involvement may also enhance children's behavior at home and in the classroom as parents and teachers work together to enhance social functioning and address problem behaviors. A growing literature has demonstrated benefits of parent involvement for social functioning (Izzo et al., 1999; McWayne et al., 2004; Reynolds, 1989; Rimm-Kaufman, Pianta, Cox, & Bradley, 2003; Supplee, Shaw, Hailstones, & Hartman, 2004). For example, a recent study of Head Start students revealed that parent involvement was associated with lower conduct problems (Fantuzzo et al., 2004). Such findings are also evident in adolescence (Hill et al., 2004). However, studies addressing parent involvement's links to socioemotional skills have typically focused on early childhood and utilized cross-sectional designs.

Research Questions

The aim of the current study is to extend past research by examining within- and between-child associations among parent involvement and academic and socioemotional trajectories during elementary school. In doing so, this investigation addresses the heavy reliance on contemporaneous, within-grade research during early childhood in the past literature by examining associations between parent involvement and academic and socioemotional functioning longitudinally throughout middle childhood.

Analytic Strategy

A major challenge facing researchers is determining whether parent involvement itself is actually beneficial to children or whether parents who are more engaged and involved in children's schooling also differ in key selection characteristics (e.g., motivation, cognitive competence) from less involved parents. If the most competent parents with developmentally advanced children are more likely to be involved in school, this could lead to an upward bias of the effect of parent involvement, thereby making it appear to be disproportionately beneficial to children. In contrast, if parents of the most struggling children are more likely to be involved, a downward bias in the influence of parent involvement may emerge, indicating that parent involvement is less beneficial for children than is actually the case. Between-child analyses are susceptible to these sorts of endogeneity or selection biases (Duncan, Magnuson, & Ludwig, 2004). One approach for addressing these biases is conducting within-individual analyses linking changes in parent involvement to changes in academic or behavioral functioning (Allison, 1990).

The present study considers between- and within-child associations between parent involvement and children's development using two-level hierarchical learn models that were estimated using full information maximum likelihood techniques in HLM 6.06. (Scientific Software International, Lincolnwood, IL). Our analysis began with the estimation of unconditional growth models to establish appropriate model specifications for our conditional analytic models (Raudenbush & Bryk, 2002). Next, a series of conditional models were estimated to address the study's aims. Level 1 contained repeated measures of achievement scores and behavior ratings taken at first, third, and fifth grades, which were nested within individual children at Level 2. Parent involvement within each grade was specified as a Level 1, along with time-varying child, family, and classroom characteristics. The Level 1 model is shown in Equation 1:

$$\begin{aligned} Y_{it} = {} & \pi_{0i} + \pi_{1i}(\mathrm{AGE}_{it} - \overline{\mathrm{AGE}_{i.}}) \\ & + \pi_{2i}(\mathrm{PINV}_{it} - \overline{\mathrm{PINV}_{i.}}) + \pi_{3i}(\mathrm{FAM}_{it} - \overline{\mathrm{FAM}_{i.}}) \\ & + \pi(\mathrm{CLASS}_{it} - \overline{\mathrm{CLASS}_{i.}}) + \varepsilon_{it.} \qquad (1) \end{aligned}$$

Achievement and social skills for child i at time t were modeled as a function of time-varying measures of age (AGE_{it}) and parent involvement (PINV_{it}), as well as family (FAM_{it}) and classroom (CLASS_{it}) characteristics. All of the predictors at Level 1 were group-mean centered, also known as within-person centering (Raudenbush & Bryk, 2002; Singer & Willett, 2003). The intercept represents each individual's average level of achievement or social skills during middle childhood, while the coefficient on AGE_{it} represents growth in achievement or social skills per month increase in age from first through fifth grade. Coefficients on the Level 1 predictors represent within-child associations over time and test whether within-child changes in parent involvement promote changes in achievement and socioemotional development. Group-mean centering is an important technique for addressing bias due to unobserved heterogeneity or unmeasured factors that vary across individuals and have a consistent effect over time. However, that these predictors are still susceptible to time-varying omitted variable bias. Between-child associations between average parent involvement and children's average level and growth of academic and socioemotional functioning were considered at Level 2. Variability in the intercept and the linear growth terms in the Level 1 equations were modeled using Equations 2 and 3:

$$\begin{aligned}\pi_{0i} = \beta_{00} &+ \beta_{01}(\text{AvPINV}_i - \overline{\text{AvPINV}}..)\\ &+ \beta_{02}(\text{EFAM}_i - \overline{\text{EFAM}}..)\\ &+ \beta_{03}(\text{CHILD}_i - \overline{\text{CHILD}}..)\\ &+ \beta_{04}(\text{AvFAM}_i - \overline{\text{AvFAM}}..)\\ &+ \beta_{05}(\text{AvCLASS}_i - \overline{\text{AvCLASS}}..) + r_{0it} \quad (2)\end{aligned}$$

$$\begin{aligned}\pi_{1i} = \beta_{10} &+ \beta_{11}(\text{AvPINV}_i - \overline{\text{AvPINV}}..)\\ &+ \beta_{12}(\text{EFAM}_i - \overline{\text{EFAM}}..)\\ &+ \beta_{13}(\text{CHILD}_i - \overline{\text{CHILD}}..)\\ &+ \beta_{14}(\text{AvFAM}_i - \overline{\text{AvFAM}}..)\\ &+ \beta_{15}(\text{AvCLASS}_i - \overline{\text{AvCLASS}}..) + r_{1it}. \quad (3)\end{aligned}$$

Variation in mean levels of achievement and social skills (π_{0i}) and in the growth of performance on these measures over time (π_{1i}) was explained with individual-level averages from middle childhood for parent involvement (AvPINV_i), family (AvFAM_i) and classroom (AvCLASS_i) characteristics. Time-invariant measures of early family and demographic (EFAM_i) characteristics from birth through 54 months, as well as time-invariant child characteristics and 54-month scores on each of the child outcomes (CHILD_i) were included as covariates. In the Level 2 equations, all predictors were grand-mean centered. These models capture both between- and within-child variation. The Level 1 coefficient on parent involvement examines whether within-child changes in parent involvement are associated with improvements in academic and social development. The Level 2 coefficients on parent involvement consider whether children with parents who have higher average levels of involvement in school display better average levels of academic and social skills as well as greater improvement in these domains over time.

Method

Sample

Data for this study are drawn from the National Institute of Child Health and Human Development Study of Early Child Care and Youth Development (NICHD SECCYD), an ongoing longitudinal, multimethod study of 1,364 children and their primary caregivers from 10 U.S. data collection sites (NICHD Early Child Care Research Network [ECCRN], 1993). Sampling for the NICHD SEYCCD was conditionally random and excluded mothers younger than 18 years of age at the time of the participating child's birth, families not intending to remain within proximity of the data collection sites for at least 3 years, children with obvious disabilities at birth or who remained in the hospital more than 7 days after birth, and mothers who would have significant problems conversing in English (NICHD ECCRN, 2004). The NICHD SECCYD contains a wide range of academic and socioemotional skills from observational and standardized assessments and parent and teacher reports, as well as parent and teacher reports of parent involvement in elementary school.

Data from birth (1991) to fifth grade were utilized in the present study. Although the initial sample comprised 1,364 children, as described in past research, ethnic minority families with lower educational levels were less likely to remain in the study over time and 137 children were lost within the first phase (1–36 months) to selective attrition (Duncan & Gibson-Davis, 2006; NICHD ECCRN, 2003). By first grade, an additional 94 children withdrew from the study, resulting in 231 attrition cases. Independent samples *t* tests and chi-square tests were conducted to examine whether the children and families in the attrition group ($n = 231$) were systematically different from the remaining sample who continued to participate during elementary school ($n = 1{,}133$). Results revealed few differences between these two groups. The retained and attrition samples did not significantly differ in maternal age, maternal cognitive ability, or child sex. However, the retained sample comprised fewer non-Hispanic Black children (11.6%; $\chi^2 = 6.44$, $p < .05$) than the attrition sample (17.7%). Mothers in this sample had more years of education ($M = 14.4$; $\chi^2 = 55.08$, $p < .001$) and were more likely to have been married when their child was born (78.6%; $\chi^2 = 349.97$, $p < .001$) than mothers in the attrition sample ($M = 13.3$ and 66.7%, respectively). Income-to-needs ratios were significantly higher among families in the retained ($M = 3.60$) than attrition ($M = 2.94$; $t = 2.76$, $p < .01$) samples.

Traditional approaches to handling missing data, such as listwise deletion or mean imputation, have been criticized for biasing estimates, misrepresenting statistical power, and leading to invalid conclusions (Acock, 2005; Rubin, 1987; Widaman, 2006). Missing data patterns were carefully analyzed and Little's MCAR tests indicated that the missing data were not missing completely at random (MCAR; $\chi^2 = 32{,}458.01$, $p < .001$), which supported the use of missing data imputation (Little, 1988). Missing data were imputed using multiple imputations by chained equations, which was implemented in Stata 10 (Royston, 2005). Currently, there is disagreement in the literature about whether missing data on dependent variables should be imputed. Some argue that imputed values for dependent variables should be deleted postimputation (von Hippel, 2007), while others maintain that the nature of missing values for independent and dependent variables are fundamentally equal and should be treated accordingly (Allison, 2008; Schafer & Graham, 2002). In the present study, imputations were performed on independent and dependent variables and imputed dependent variables were retained after imputation. Multiple imputations were performed in Stata, and based on the relative efficiency calculation by Rubin (1987), five imputations were deemed sufficient for the level of missing data in our study, resulting in a final sample of 1,364 children. The analyses for this study were performed using HLM 6.06 (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2004). Following imputation, the five data sets were imported into HLM, which was then used to conduct separate analyses for each set of plausible values. Parameter estimates were averaged and standard errors were computed in HLM 6.06 using standard techniques described by Raudenbush et al. (2004).

Measures and Procedures

Parent Involvement

Information on parent involvement was obtained from parents and teachers in first, third, and fifth grades using modified versions of the Parent–Teacher Involvement Questionnaire (Miller-Johnson, Maumary-Gremaud, & the Conduct Disorders Research Group, 1995). These questionnaires included items regarding the frequency and the quality of parents' involvement in children's educational progress in school and at home. A 5-point Likert-type response scale was employed with responses ranging from (1) *not at all* to (5) *a great deal*. Our composite measures of parent involvement include items reflecting teachers' and parents' reports of parental encouragement of education (e.g., ''How often does this parent volunteer or visit at school?''), parental investment (e.g., ''How important is education in this family?''), and educational attitudes (e.g., ''How much do you feel this parent has the same goals for his or her child that the school does?''). Parent and teacher parent involvement composites comprised 12 items and 10 items, respectively, with higher scores indicating more parent involvement. In addition, the parent involvement composites were highly reliable within each first, third, and fifth grades (parents' Cronbach's αs = .85–.86; teachers' Cronbach's αs = .89–.93). It is important to note that items selected to create the mother and teacher composites were identical for each reporter across the first, third, and fifth grade years. Parents' and teachers' reports were both analyzed in separate models rather than relying on a single reporter (see Table 1). Average levels of parent involvement were comparable across grades (see Table 1).

Outcomes

Academic achievement. Academic achievement was assessed at 54 months (prior to kindergarten entry), and first, third, and fifth grades using three subtests from the Woodcock–Johnson Psycho-Educational Battery–Revised (WJ–R; Woodcock & Johnson, 1989, 1990), which was administered to participants in a laboratory setting. Early childhood scores were included as an additional preventative measure against omitted variable bias.

Picture Vocabulary tested comprehension, knowledge and receptive vocabulary, Letter–Word Identification assessed symbolic learning and reading skills, and Applied Problems tested mathematical skills. The scores used in the current study's analyses were *W* scores, a special transformation of the Rasch ability scale, which ease interpretation of performance and facilitate analyses of change over time. Generally, high internal consistency estimates have been found in standardized samples (Woodcock & Johnson, 1989, 1990). Internal consistencies have been reported ranging from .70 to .82 for the Picture Vocabulary test with test–retest reliabilities ranging from .63 to 0.78. Internal consistencies for Letter–Word Identification and Applied problems have been quite high, ranging from .94 to .98 with test–retest reliabilities of .80–.87 (Woodcock & Johnson, 1989, 1990).

Socioemotional development. Socioemotional development was assessed during first, third, and fifth grades using two standardized measures. Parents completed the Child Behavior Checklist (CBCL; Achenbach, 1991a) and teachers completed the

Table 1
Descriptive Statistics

	Early childhood composites[a]		First grade		Third grade		Fifth grade	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Maternal report of parent involvement			4.63	0.50	4.62	0.41	4.56	0.49
Teacher report of parent involvement			3.73	0.90	4.22	0.64	3.55	0.92
Income-to-needs ratio	3.48	2.86	3.76	3.00	4.22	3.67	4.38	3.94
Children in the household			2.38	0.95	2.41	0.98	2.48	1.04
Mother married			0.69		0.73		0.73	
Maternal employment	23.04	19.13	26.50	19.09	26.84	19.11	27.67	19.03
Teacher experience			12.21	10.77	11.87	10.55	14.50	9.46
Classroom quality			5.32	1.29	5.05	0.78	5.10	0.71
Teacher-reported SSRS			103.14	13.67	102.30	14.51	103.36	14.31
Teacher-reported CBCL			49.91	9.43	51.43	9.71	50.29	9.75
Mother-reported SSRS	98.32	13.59	105.07	15.07	105.98	15.95	107.38	15.07
Mother-reported CBCL	51.11	9.54	48.52	9.97	47.69	10.40	46.37	10.74
Math	423.87	19.46	469.42	15.67	497.16	12.63	509.77	12.29
Reading	368.70	21.39	452.05	24.06	493.39	18.27	509.88	17.38
Vocabulary	459.21	13.71	483.42	12.28	496.50	11.42	505.52	11.90
Child male	0.48							
Child Black	0.13							
Child Hispanic	0.06							
Child other ethnicity	0.05							
Maternal age	28.11	5.63						
Maternal education	14.23	2.51						
Maternal PPVT score	98.11	18.60						
Average HOME scores	−0.04	0.87						
Middle childhood average family variables[b]								
Mother married	0.72							
Maternal employment	27.00	16.44						
Income-to-needs ratio	4.12	3.35						
Children in the household	2.42	0.94						
Teacher experience	12.86	6.44						
Classroom quality	34.43	3.07						

Note. N = 1,364. SSRS = Social Skills Rating System; CBCL = Child Behavior Checklist; PPVT = Peabody Picture Vocabulary Test; HOME = Home Observation for Measure of the Environment.
[a]Data treated as time invariant. [b]Data averaged across first, third, and fifth grades.

Teacher Report Form (TRF), a slightly modified version of the CBCL, to measure children's social competence and behavior problems (Achenbach, 1991a). As with the achievement outcomes, early ratings from the 54-month time of social skills and problem behaviors were included in the between-child analyses when predicting later problem behaviors. Reporters rated how true the behaviors are to a child on a 3-point scale (0 = *not true*, 1 = *sometimes or somewhat true*, 2 = *very true or often true*). Higher overall scores reflect more social or behavioral problems. In the present analyses, a total behavior problem standardized t score, comprising two subscale scores of externalizing and internalizing problems, was utilized. Externalizing problems include aggressive or delinquent behavior, and internalizing problems comprised withdrawn, anxious, or depressed behaviors. The CBCL and TRF are widely accepted as reliable and internally consistent (Achenbach, 1991b). Parents' and teachers' reports of problem behaviors were highly reliable at each time point (parents' Cronbach's αs: = .93–.95; teachers' Cronbach's αs: = .95–.96).

In first, third, and fifth grades, parents and teachers also completed the Social Skills Rating System (SSRS; Gresham & Elliott, 1990), which comprised two subscales examining children's social skills and academic competence. Parents also completed the SSRS at 54 months. This score was used to control for social skills prior to school entry in the between-child analyses. The current study focused solely on the social skills subscale which assesses cooperation, assertion and self-control. Cooperation includes behaviors such as paying attention to the teacher's instruction and putting away work materials properly. Assertion comprised initiating behaviors, such as starting conversations with peers and helping peers with classroom tasks. Self-control includes behaviors that emerge in conflict situations, such as responding to teasing or peer pressure appropriately, receiving criticisms well, and controlling temper. Parents and teachers assessed the frequency of a social behavior on a 3-point scale (0 = *never*, 1 = *sometimes*, 2 = *very often*). Additionally, the reporters rated how important the behavior is for classroom success, again on a 3-point scale (0 = *not important*, 2 = *important*, 3 = *critical*). The scores from all three of these subscales were summed to create a total SSRS score for each child. These scores were standardized with higher scores reflecting better social adjustment. The validity of the SSRS is well documented and the scale has also been found to be highly correlated with other well established measures of social functioning such as the CBCL (Achenbach, 1991a, 1991b; Gresham & Elliott, 1990). Parents' and teachers' reports of social skills were highly reliable at each time point (parents' Cronbach's αs = .92–.94; teachers' Cronbach's αs = .93–.94).

Covariates

Child characteristics. A number of important background characteristics were included as covariates of child outcomes and parent involvement (see Table 1). Child demographic information was obtained from periodic parent interviews from the child's 1st month of life through fifth grade. At 1 month, mothers reported on child sex (*0 = female, 1 = male*) and ethnicity (non-Hispanic White, Hispanic, non-Hispanic Black, and Other).

Classroom characteristics. Teacher ratings were included to measure parent involvement, children's social skills and their behavior problems. Thus, it was important to control for background characteristics of teachers, such as teaching experience as well as general teacher practices that may influence their ratings or children's academic or social progress throughout school. In the fall of each school year, first-, third-, and fifth-grade teachers reported on their years of teaching experience in a Teacher Questionnaire that was modified from the Schools and Staffing Survey (National Center for Educational Statistics, 1994). Across elementary school, teachers had comparable levels of teaching experience (see Table 1).

Additionally, teacher practices were observed to measure the classroom environment as a control for classroom quality. Scores for positive emotional climate in the classroom were obtained from the qualitative rating portion of the Classroom Observation System (COS; NICHD ECCRN, 2002). This scale was chosen due to its consistent use across first, third, and fifth grades. Ratings were obtained during 2-hr classroom observations of which 30 min were dedicated to qualitative ratings. These 30 min were divided into three 10-min qualitative segments during which coders rated the teacher, the child and the classroom. Scores on the ratings used in the current study ranged from 1 to 7 and represented a global impression of the overall enthusiasm, warmth and positive regard expressed by the teacher and students in the classroom. First-, third-, and fifth-grade ratings were entered in the Level 1 models and also averaged into a single covariate and included in the Level 2 models. Classroom quality scores were comparable across first ($M = 5.32$, $SD = 1.29$), third ($M = 5.05$, $SD = 0.78$), and fifth ($M = 5.10$, $SD = 0.71$) grades.

Family characteristics. Family demographic information was also obtained from parent interviews from the child's birth through fifth grade. These characteristics include maternal age, the number of years of education completed before the birth of the study child, and the number of children in the household in first, third, and fifth grades. Mothers also reported on their marital status (0 = *married*, 1 = *unmarried*) and hours of employment at first, third, and fifth grades.

Income data were frequently collected from families participating in the NICHD SECCYD. Income-to-needs ratios were calculated by taking pretax household income for a given year and dividing it by the poverty threshold for a household, which takes into account the year the income was earned, the number of individuals in the home, and the number of children in the home. Income-to-needs ratios from first, third, and fifth grades were used in estimating within-child effects of parent involvement on achievement and social skills. A composite variable of the average income-to-needs ratio across the three grades was also used in estimating between-child effects of parent involvement on levels and changes in children's academic and social skills. An average early income-to-needs ratio composite was computed using income data from

6, 15, 24, 36, and 54 months as another covariate in the between-child analysis.

Maternal scores on the Peabody Picture Vocabulary Test–Revised (PPVT–R, Dunn & Dunn, 1981) from a laboratory assessment at the 36-month time point were used to assess maternal cognitive competence. Split-half correlations for reliability ranged from .80 to .83 in a standardization sample (Dunn & Dunn, 1981). Since maternal cognitive ability has the potential to influence child academic and socioemotional outcomes, and therefore confound associations between parent involvement and these outcomes, the PPVT–R was used as time-invariant covariate in the Level 2 models.

The Home Observation for Measure of the Environment (HOME; Caldwell & Bradley, 1984) was used to represent each child's early home environment. The HOME is a widely used assessment tool which integrates investigator observation and parental interview components to examine the physical environment, responsiveness, and academically stimulating characteristics of a home. Composite scores from age-specific versions of the HOME were obtained at 15, 36, and 54 months. These scores were transformed into z scores and then averaged to create a single early childhood home environment composite.

Results

Descriptive statistics regarding the sample and outcomes are listed in Table 1. Females comprised 52% of the current sample and 78% of the children were non-Hispanic White. On average, mothers were in their late 20s ($M = 28.11$, ages ranged from 18 to 46) with some college education ($M = 14.23$, values ranged from 7 to 21 years) and most mothers reported being stably married between the first and fifth grades ($M = 0.72$, values ranged from 0 to 1). The mean income-to-needs ratio was generally high and increased during school.

As shown in Table 2, achievement scores improved significantly across elementary school from first to fifth grade between children, as well as parent SSRS and CBCL ratings. Teacher ratings on the CBCL ($M = 50.54$, values ranged from 49.91 to 51.43) and their ratings on the SSRS ($M = 102.93$, values ranged from 102.30 to 103.36) did not change significantly across grades. On average, parent involvement was rated comparably across grades for parent reports ($M = 4.56$–4.63). Teachers rated parent involvement the highest in the third grade ($M = 3.74$ vs. 4.22 vs. 3.55, respectively). Mean comparisons using simple t tests revealed that parents rated parent involvement higher than teachers at each grade ($p < .001$).

We began our analysis by estimating unconditional growth models, which revealed significant variability in the intercepts and slopes of the trajectories for our outcome variables (see Table 2). The only exception was that the age parameter for teachers' reports of social skills and behavior problems demonstrated low reliability and chi-square tests showed no variability in this parameter, so we fixed the linear growth terms in the Level 2 models predicting teachers' reports of social skills and behavior problems. Since all of the other outcomes demonstrated significant reliability of and variability on the age parameter, both their intercepts and

Table 2

Unconditional Growth Models for Achievement and Social Skills Across Elementary School

	Individual growth curve parameters		Individual growth curve variance components		
	Average score[a] (*SE*)	Age[b] (*SE*)	Residual variance	Intercept variance (*df* = 1039)	Slope variance (*df* = 1039)
Reading	485.02*** (0.57)	1.20*** (0.01)	152.59	293.88***	0.02
Math	492.09*** (0.38)	0.84*** (0.01)	72.15	125.17***	0.00
Vocabulary	495.17*** (0.30)	0.46*** (0.01)	33.33	104.00***	0.01***
Teacher-reported SSRS[c]	102.93*** (0.35)	0.01 (0.01)	124.25	76.60***	
Mother-reported SSRS	106.14*** (0.39)	−0.05*** (0.01)	63.41	166.86***	0.02***
Teacher-reported CBCL[c]	50.54*** (0.22)	0.01 (0.01)	53.92	39.21***	
Mother-reported CBCL	47.53*** (0.27)	−0.04*** (0.00)	25.25	79.81***	0.01***

Note. Significance levels denoted on the intercept and slope variance components reflect whether there were significant individual differences across individual intercepts and slopes. SSRS = Social Skills Rating System; CBCL = Child Behavior Checklist.
[a]Intercept. [b]Slope, rate of change across first, third, and fifth grades. [c]Nonrandom Level 1 coefficient.
***$p < .001$.

Table 3
Multilevel Growth Models of Maternal Reports of Parent Involvement Predicting Achievement

	Reading^{a}		Math^{a}		Vocab^{a}	
	Intercept (*SE*)	Slope (*SE*)	Intercept (*SE*)	Slope (*SE*)	Intercept (*SE*)	Slope (*SE*)
Level 1						
Parent involvement	—	−0.48 (0.75)	—	−0.05 (0.73)	—	0.35 (0.48)
Income-to-needs ratio	—	−0.14 (0.15)	—	0.08 (0.13)	—	0.07 (0.11)
Children in the household	—	−0.34 (0.83)	—	−0.21 (0.64)	—	0.70 (0.45)
Maternal marital status	—	−0.82 (1.19)	—	−0.48 (1.17)	—	−0.16 (0.81)
Hours of maternal employment	—	0.01 (0.02)	—	−0.02 (0.02)	—	−0.00 (0.02)
Teacher experience	—	0.00 (0.02)	—	−0.01 (0.02)	—	0.03* (0.01)
Classroom quality	—	−0.92*** (0.04)	—	−0.52*** (0.03)	—	−0.13*** (0.02)
Level 2						
Intercept	485.10*** (0.47)	1.11*** (0.01)	492.12*** (0.30)	0.79*** (0.01)	495.15*** (0.20)	0.44*** (0.01)
Average parent involvement^{a}	−0.34 (1.57)	0.04 (0.05)	−1.13 (1.00)	0.01 (0.03)	0.15 (0.77)	−0.01 (0.03)
Early childhood family variables						
Child male	−0.06 (0.94)	−0.04 (0.03)	−3.06*** (0.53)	0.05** (0.01)	−1.19* (0.51)	0.03** (0.01)
Child Black	−3.33 (1.84)	−0.02 (0.05)	−1.89 (1.00)	−0.01 (0.03)	−0.79 (0.80)	−0.04* (0.02)
Child Hispanic	2.33 (1.87)	0.06 (0.06)	−0.27 (1.07)	0.06 (0.03)	1.76* (0.79)	−0.00 (0.03)
Child other ethnicity	−1.76 (2.22)	0.05 (0.06)	1.03 (1.34)	0.06 (0.04)	3.16** (1.05)	−0.03 (0.03)
Maternal age	−0.03 (0.12)	0.00 (0.00)	0.04 (0.06)	−0.00 (0.00)	−0.04 (0.05)	−0.00 (0.00)
Maternal education	0.23 (0.24)	−0.00 (0.01)	0.20 (0.18)	−0.00 (0.00)	0.21 (0.15)	0.00 (0.00)
PPVT score^{b}	0.15** (0.04)	0.00 (0.00)	0.05* (0.02)	0.00 (0.00)	0.15*** (0.02)	0.00 (0.00)
Average early income-to-needs^{c}	−0.17 (0.24)	0.01 (0.01)	0.09 (0.15)	0.00 (0.00)	0.15 (0.15)	0.00 (0.01)
Average HOME scores^{d}	1.20 (0.86)	0.03 (0.03)	1.06* (0.49)	0.02 (0.02)	1.02 (0.48)	−0.01 (0.01)
Early achievement score^{e}	0.40*** (0.02)	−0.01*** (0.00)	0.34*** (0.02)	−0.00*** (0.00)	0.38*** (0.02)	−0.00** (0.00)
Middle childhood family variables^{a}						
Average marital status	2.39* (1.17)	−0.05 (0.04)	0.80 (0.75)	0.04 (0.03)	0.39 (0.76)	−0.00 (0.02)
Average maternal employment	0.02 (0.03)	−0.00 (0.00)	0.01 (0.02)	0.00 (0.00)	−0.00 (0.02)	0.00 (0.00)
Average childhood income	0.19 (0.25)	−0.00 (0.01)	0.17 (0.16)	−0.00 (0.00)	−0.16 (0.13)	−0.00 (0.00)
Average children in home	−0.45 (0.62)	−0.01 (0.02)	−0.17 (0.33)	.01 (0.01)	−0.88** (0.27)	−0.01 (0.01)
Classroom variables^{a}						
Average teacher experience	0.01 (0.06)	0.00 (0.00)	0.02 (0.04)	0.00 (0.00)	0.03 (0.03)	0.00 (0.00)
Average classroom quality	−0.07 (0.16)	−0.02*** (0.00)	−0.01 (0.11)	−0.01*** (0.00)	0.08 (0.09)	−0.01* (0.00)

Note. $N = 1,364$. PPVT = Peabody Picture Vocabulary Test; HOME = Home Observation for Measure of the Environment.
^{a}Data from first, third, and fifth grades. ^{b}Data from 36 months of age. ^{c}Data from 6 to 54 months of age. ^{d}Data from 15 to 54 months of age. ^{e}Data from 54 months of age.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

linear growth terms were estimated as random effects at Level 2. As the slope terms suggest, children's reading, math, vocabulary skills, as well as their parent-rated social skills increased over time, while their problem behaviors declined.

Predicting Academic Trajectories

Within-Child Effects

Tables 3 and 4 present results of the HLM analyses predicting achievement over time with mother or teacher reports of parent involvement. The first panel of Table 3 shows the coefficients from the Level 1 model, which assessed whether within-person changes in maternal reports of parent involvement predicted within-child changes in reading, math and vocabulary scores during elementary school, controlling for changes in other child, family and school characteristics as well as between-person differences in all of these factors. Table 4 displays the same set of results for models using teacher reports of parent involvement. Increases in parent involvement were largely unrelated to individual growth in academic skills, regardless of whether parents or teachers reported on parent involvement practices. One exception emerged for reading skills. Increases in teacher-reported parent involvement were related to declines in children's reading scores (see Table 4).

Table 4

Multilevel Growth Models of Teacher Reports of Parent Involvement Predicting Achievement

	Reading[a]		Math[a]		Vocab[a]	
	Intercept (*SE*)	Slope (*SE*)	Intercept (*SE*)	Slope (*SE*)	Intercept (*SE*)	Slope (*SE*)
Level 1						
Parent involvement	—	−0.75* (0.32)	—	−0.39 (0.28)	—	−0.64 (0.30)
Income-to-needs ratio	—	−0.14 (0.15)	—	0.07 (0.13)	—	0.06 (0.11)
Children in the household	—	−0.31 (0.81)	—	−0.20 (0.63)	—	0.73 (0.45)
Maternal marital status	—	−0.75 (1.20)	—	−0.49 (1.19)	—	−0.17 (0.83)
Hours of maternal employment	—	0.01 (0.02)	—	−0.02 (0.02)	—	−0.00 (0.02)
Teacher experience	—	0.00 (0.02)	—	−0.01 (0.02)	—	0.03* (0.01)
Classroom quality	—	−0.91*** (0.04)	—	−0.52*** (0.03)	—	−0.13*** (0.02)
Level 2						
Intercept	485.10*** (0.46)	1.10*** (0.01)	492.12*** (0.30)	0.78*** (0.01)	495.15*** (0.20)	0.43*** (0.01)
Average parent involvement[a]	1.97 (1.13)	−0.03 (0.03)	1.03 (0.53)	0.00 (0.02)	0.45 (0.66)	0.01 (0.01)
Early childhood family variables						
Child male	−0.02 (0.95)	−0.04 (0.03)	−0.30*** (0.52)	0.05** (0.01)	−1.18* (0.51)	0.04** (0.01)
Child Black	−3.53 (1.91)	−0.02 (0.05)	−1.98* (0.99)	−0.01 (0.03)	−0.85 (0.79)	−0.04* (0.02)
Child Hispanic	2.42 (1.87)	0.06 (0.06)	−0.15 (1.08)	0.06 (0.03)	1.76* (0.80)	0.00 (0.03)
Child other ethnicity	−1.57 (2.21)	0.05 (0.05)	1.16 (1.32)	0.06 (0.04)	3.19** (1.05)	−0.02 (0.03)
Maternal age	−0.05 (0.12)	0.00 (0.00)	0.03 (0.06)	−0.00 (0.00)	−0.04 (0.05)	−0.00 (0.00)
Maternal education	0.19 (0.24)	−0.00 (0.01)	0.18 (0.18)	−0.00 (0.00)	0.20 (0.14)	0.00 (0.00)
PPVT score[b]	0.15** (0.04)	0.00 (0.00)	0.05* (0.02)	0.00 (0.00)	0.15*** (0.02)	0.00* (0.00)
Average early income-to-needs[c]	−0.20 (0.24)	0.01 (0.01)	0.07 (0.15)	0.00 (0.00)	0.14 (0.15)	0.00 (0.01)
Average HOME scores[d]	0.69 (1.00)	0.04 (0.03)	0.70 (0.55)	0.02 (0.02)	0.93 (0.54)	−0.01 (0.01)
Early achievement score[e]	0.40*** (0.02)	−0.01*** (0.00)	0.34*** (0.02)	−0.00*** (0.00)	0.38*** (0.02)	−0.00** (0.00)
Middle childhood family variables[a]						
Average marital status	1.98 (1.18)	−0.04 (0.04)	0.54 (0.75)	0.04 (0.03)	0.31 (0.74)	−0.01 (0.02)
Average maternal employment	0.04 (0.03)	−0.00 (0.00)	0.02 (0.02)	0.00 (0.00)	−0.00 (0.02)	0.00 (0.00)
Average childhood income	0.19 (0.24)	−0.00 (0.01)	0.17 (0.16)	−0.00 (0.00)	−0.15 (0.13)	−0.00 (0.00)
Average children in home	−0.36 (0.62)	−0.01 (0.02)	−0.13 (0.32)	0.01 (0.01)	−0.86** (0.26)	−0.01 (0.01)
Classroom variables[a]						
Average teacher experience	0.01 (0.06)	0.00 (0.00)	0.02 (0.04)	0.00 (0.00)	0.02 (0.03)	0.00 (0.00)
Average classroom quality	−0.07 (0.17)	−0.02** (0.01)	−0.02 (0.11)	−0.01*** (0.00)	0.08 (0.09)	−0.01* (0.00)

Note. $N = 1{,}364$. PPVT = Peabody Picture Vocabulary Test; HOME = Home Observation for Measure of the Environment.

[a] Data from first, third, and fifth grades. [b] Data from 36 months of age. [c] Data from 6 to 54 months of age. [d] Data from 15 to 54 months of age. [e] Data from 54 months of age.

$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

More specifically, a standard deviation improvement in teacher-reported parent involvement was associated with a .03 *SD* decline in reading skills. The only other variable that consistently, significantly predicted within-child changes in achievement was classroom quality. Increases in the emotional quality of a child's classroom were associated with small declines (ranging from .01 to .05 *SD*) in each of the achievement domains.

Between-Child Effects

The second panel of Table 3 shows results from the Level 2 models that consider between-child associations of average levels of parent involvement with average level of achievement (intercept) and achievement growth over time (slope). Parent-reported parent involvement, when averaged across elementary school, was not significantly associated with children's average reading, math, and vocabulary achievement scores. Results also showed that there were no significant associations between average levels of parent-reported parent involvement and growth in achievement over time. The second panel of Table 4 shows a similar pattern emerged when teachers reported on parent involvement. Overall, few significant relations emerged between parent involvement and achievement; specifically, no between-child associations were detected and small within-child associations were evident only for reading skills when teachers reported on parent involvement.

In contrast, several child, family and classroom characteristics were linked to between-child differences in achievement, and similar patterns were revealed for both maternal and teacher reports of parent involvement. Maternal cognitive competence was associated with higher average levels of reading, math and vocabulary. Male children had lower average achievement scores in math and vocabulary than females, but no gender differences were detected for reading. Earlier scores from the 54-month time point on each of the achievement tests were positively related to average level scores in elementary school. In regards to the growth of academic skills over time, children in less emotionally supportive classes experienced greater achievement gains over time.

Predicting Social Skill Trajectories

Within-Child Effects

Table 5 presents results of the HLM analysis predicting children's social skills over time with maternal reports of parent involvement. The first panel of Table 5 displays the coefficients from the Level 1 model, within-child changes in mother-reported parent involvement predicting within-child changes in parent- and teacher-reported social skills and problem behaviors during elementary school. Within-child increases in parent involvement across elementary school were related to improvements in teacher-reported social skills. A standard deviation increase in parent involvement was linked to a modest increase of 0.22 *SD* in teacher-reported social skills, and small reductions in problem behaviors over time: 0.12 *SD* in teacher-rated problem behaviors and 0.08 *SD* mother-rated problem behaviors. In contrast, within-child changes in maternal reports of parent involvement were not linked to changes in maternal reports of child social skills.

Table 6 presents results of the HLM analysis predicting social skills over time with teachers' reports of parent involvement. The first panel of Table 6 displays the coefficients from the Level 1 model. Within-child increases in teacher-reported parent involvement were related to 0.12 *SD* improvement in teacher-reported social skills and 0.08 *SD* decrease in teacher-reported behavior problems. In contrast, enhancements in teacher-reported parent involvement were not associated with changes in maternal reports of child social skills or behavior problems.

Between-Child Effects

The second panel of Table 5 shows results from the Level 2 model examining between-child associations of parent-reported parent involvement with average levels of social skills (intercept) and behavior problems as well as growth in social skills and behavior problem over time (slope). Higher average levels of maternal-reported parent involvement across elementary school were related to better social functioning across parent and teacher SSRS reports. When compared to children whose parents' involvement in school was at the sample average, children whose parents scored a standard deviation above average displayed social skills levels that were 0.09 and 0.15 *SD* higher based on teacher and mother reports, respectively. In contrast, higher levels of maternal-reported parent involvement were not significantly associated with maternal- or teacher-rated problem behaviors on the CBCL. In summary, between-individual differences in average parent involvement across elementary school predicted higher average social functioning but not problem behaviors problems. Finally, average levels of maternal-reported parent involvement did not predict growth in social skills and problem behaviors over time.

As with the achievement findings, several child, family, and school characteristics were also linked to between-child differences in social skills and problem behaviors when parent involvement was reported by parents. Earlier ratings of social skills and problem behaviors from the 54-month time point were associated with average level scores in their corresponding subsequent ratings except for teacher-reported social skills. Higher classroom quality and teacher experience promoted higher average levels of teacher-rated social skills and were linked to lower average levels in the teacher- and maternal-rated problem behaviors.

The second panel of Table 6 show results from the Level 2 model examining between-child associations between teacher-reported parent involvement and trajectories of social skills and problem behaviors over time. Higher average levels of teacher-reported parent involvement were related to better social functioning and lower rates of behavior problems based on parent- and teacher reports. When compared to children whose parents' involvement in school was at the sample average, children whose parents scored a standard

Table 5

Multilevel Growth Models of Maternal Reports of Parent Involvement Predicting Social Skills and Problem Behaviors

	Teacher-reported SSRS		Mother-reported SSRS		Teacher-reported CBCL		Mother-reported CBCL	
	Intercept (*SE*)	Slope (*SE*)	Intercept (*SE*)	Slope (*SE*)	Intercept (*SE*)	Slope (*SE*)	Intercept (*SE*)	Slope (*SE*)
Level 1								
Parent involvement	—	3.37*** (0.64)	—	−0.32 (0.54)	—	−1.25** (0.38)	—	−0.92* (0.38)
Income-to-needs ratio	—	−0.24 (0.19)	—	0.14 (0.17)	—	0.08 (0.13)	—	−0.03 (0.11)
Children in the household	—	.92 (0.85)	—	1.05 (0.63)	—	−0.65 (0.53)	—	0.26 (0.34)
Maternal marital status	—	−0.58 (1.34)	—	1.07 (0.98)	—	0.29 (0.71)	—	−1.16 (0.63)
Hours of maternal employment	—	−0.02 (0.03)	—	0.03 (0.02)	—	0.01 (0.02)	—	0.01 (0.01)
Teacher experience	—	−0.01 (0.02)	—	−0.02 (0.02)	—	0.01 (0.01)	—	0.02 (0.01)
Classroom quality	—	.11** (0.03)	—	0.03 (0.03)	—	−0.09*** (0.02)	—	−0.03 (0.02)
Level 2								
Intercept	102.93*** (0.32)	0.02* (0.01)	106.14*** (0.30)	−0.05*** (0.01)	50.54*** (0.20)	−0.00 (0.01)	47.53*** (0.20)	−0.05*** (0.01)
Average parent involvement[a]	2.46* (1.07)	−0.00 (0.04)	4.41** (1.28)	0.01 (0.03)	−0.95 (0.75)	−0.01 (0.02)	−1.30 (0.92)	0.00 (0.02)
Early childhood family variables								
Child male	−0.33 (0.65)	0.00 (0.02)	0.44 (0.66)	−0.02 (0.02)	−0.32 (0.50)	−0.02 (0.01)	−0.21 (0.37)	0.00 (0.01)
Child Black	−1.45 (1.03)	−0.02 (0.05)	−2.14 (1.25)	0.06 (0.04)	0.70 (0.92)	0.03 (0.03)	−1.00 (0.94)	−0.02 (0.02)
Child Hispanic	−0.16 (1.44)	−0.07 (0.06)	−0.63 (1.55)	0.01 (0.04)	−0.15 (0.87)	0.03 (0.03)	−1.64 (0.90)	−0.03 (0.02)
Child other ethnicity	0.51 (1.55)	−0.01 (0.06)	−0.32 (1.46)	−0.04 (0.04)	−0.87 (1.02)	−0.02 (0.03)	−1.42 (1.22)	−0.03 (0.02)
Maternal age	−0.05 (0.07)	0.00.00	0.01 (0.07)	−0.00 (0.00)	0.02 (0.05)	−0.00 (0.00)	−0.09 (0.04)	−0.00 (0.00)
Maternal education	0.32 (0.18)	−0.00 (0.01)	0.09 (0.19)	−0.00 (0.01)	−0.05 (0.15)	0.00 (0.00)	0.06 (0.12)	−0.00 (0.00)
PPVT score[b]	0.02 (0.02)	0.00 (0.00)	0.07* (0.03)	−0.00 (0.00)	−0.02 (0.02)	−0.00 (0.00)	0.02 (0.02)	0.00 (0.00)
Average early income-to-needs[c]	0.04 (0.17)	0.02* (0.01)	−0.51* (0.20)	0.01* (0.00)	−0.10 (0.14)	−0.01 (0.01)	−0.01 (0.11)	0.00 (0.00)
Average HOME scores[d]	2.98*** (0.60)	0.00 (0.02)	1.14 (0.67)	0.02 (0.01)	−2.01** (0.53)	0.01 (0.01)	−0.80* (0.38)	0.00 (0.01)
Early outcome score[e]	0.05 (0.03)	−0.00 (0.00)	0.57*** (0.03)	0.00*** (0.00)	0.13*** (0.02)	0.00 (0.00)	0.64*** (0.02)	−0.00 (0.00)
Middle childhood family variables[a]								
Average marital status	2.26* (1.04)	−0.06 (0.04)	0.65 (1.09)	−0.00 (0.02)	−1.73* (0.74)	−0.01 (0.02)	−0.68 (0.69)	0.02 (0.02)
Average maternal employment	0.02 (0.02)	0.00 (0.00)	0.03 (0.02)	−0.00 (0.00)	0.00 (0.02)	−0.00 (0.00)	−0.01 (0.01)	0.00 (0.00)
Average childhood income	0.08 (0.15)	−0.01 (0.01)	0.40* (0.16)	−0.00 (0.00)	−0.05 (0.12)	0.01 (0.00)	−0.26** (0.09)	0.00 (0.00)
Average children in home	0.57 (0.43)	0.00 (0.01)	−0.01 (0.42)	0.01 (0.01)	−0.36 (0.32)	0.00 (0.01)	−0.68** (0.24)	0.00 (0.01)
Classroom variables[a]								
Average teacher experience	0.15* (0.06)	0.00 (0.00)	−0.03 (0.05)	−0.00 (0.00)	−0.12** (0.04)	−0.00* (0.00)	0.03 (0.04)	−0.00 (0.00)
Average classroom quality	0.32** (0.11)	0.01 (0.01)	0.08 (0.13)	−0.00 (0.00)	−0.22** (0.07)	−0.00 (0.00)	−0.18* (0.08)	−0.00 (0.00)

Note. $N = 1{,}364$. SSRS = Social Skills Rating System; CBCL = Child Behavior Checklist; PPVT = Peabody Picture Vocabulary Test; HOME = Home Observation for Measure of the Environment.

[a] Data from first, third, and fifth grades. [b] Data from 36 months of age. [c] Data from 6 to 54 months of age. [d] Data from 15 to 54 months of age. [e] Data from 54 months of age.

$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Table 6

Multilevel Growth Models of Teacher Reports of Parent Involvement Predicting Social Skills and Problem Behaviors

	Teacher-reported SSRS		Mother-reported SSRS		Teacher-reported CBCL		Mother-reported CBCL	
	Intercept (*SE*)	Slope (*SE*)	Intercept (*SE*)	Slope (*SE*)	Intercept (*SE*)	Slope (*SE*)	Intercept (*SE*)	Slope (*SE*)
Level 1								
Parent involvement	—	1.83** (0.47)	—	0.14 (0.34)	—	−0.80* (0.31)	—	−0.04 (0.24)
Income-to-needs ratio	—	−0.24 (0.18)	—	0.14 (0.17)	—	0.08 (0.13)	—	−0.03 (0.10)
Children in the household	—	0.71 (0.87)	—	0.98 (0.61)	—	−0.55 (0.54)	—	0.26 (0.32)
Maternal marital status	—	−0.97 (1.33)	—	1.12 (0.96)	—	0.43 (0.72)	—	−1.05 (0.61)
Hours of maternal employment	—	−0.02 (0.03)	—	0.03 (0.02)	—	0.01 (0.02)	—	0.01 (0.01)
Teacher experience	—	−0.01 (0.03)	—	−0.02 (0.02)	—	0.00 (0.02)	—	0.02 (0.01)
Classroom quality	—	0.09* (0.04)	—	0.03 (0.03)	—	−0.08** (0.02)	—	−0.02 (0.02)
Level 2								
Intercept	102.93*** (0.31)	0.04* (0.01)	106.14*** (0.30)	−0.05*** (0.01)	50.54*** (0.19)	−0.01 (0.01)	47.53*** (0.20)	−0.05*** (0.01)
Average parent involvement[a]	6.56*** (0.62)	0.03 (0.02)	2.58*** (0.64)	.01 (0.02)	−3.81*** (0.37)	−0.01 (0.02)	−0.83 (0.54)	0.01 (0.02)
Early childhood family variables								
Child male	−0.18 (0.62)	−0.00 (0.02)	0.56 (0.67)	−0.02 (0.02)	−0.43 (0.47)	−0.02 (0.01)	−0.26 (0.37)	0.00 (0.01)
Child Black	−2.17* (0.99)	−0.03 (0.05)	−2.50 (1.30)	0.06 (0.04)	1.10 (0.88)	0.03 (0.03)	−0.86 (0.92)	−0.02 (0.02)
Child Hispanic	−0.23 (1.31)	−0.08 (0.06)	−0.97 (1.51)	.01 (0.04)	−0.17 (0.82)	0.03 (0.03)	−1.53 (0.89)	−0.03 (0.02)
Child other ethnicity	0.97 (1.56)	−0.01 (0.06)	−0.26 (1.45)	−0.04 (0.04)	−1.17 (1.01)	−0.02 (0.03)	−1.46 (1.21)	−0.03 (0.03)
Maternal age	−0.10 (0.07)	0.00 (0.00)	−0.00 (0.07)	−0.00 (0.00)	0.04 (0.04)	−0.00 (0.00)	−0.08 (0.04)	−0.00 (0.00)
Maternal education	0.16 (0.18)	−0.01 (0.01)	0.02 (0.19)	−0.00 (0.01)	0.04 (0.15)	0.00 (0.00)	0.08 (0.12)	−0.00 (0.00)
PPVT score[b]	0.02 (0.02)	0.00 (0.00)	0.07* (0.03)	−0.00 (0.00)	−0.02 (0.02)	−0.00 (0.00)	0.02 (0.02)	0.00 (0.00)
Average early income-to-needs[c]	−0.05 (0.17)	0.02* (0.01)	−0.55** (0.20)	0.01* (0.00)	−0.04 (0.14)	−0.01 (0.01)	0.00 (0.11)	0.00 (0.00)
Average HOME scores[d]	1.66* (0.61)	−0.01 (0.02)	0.97 (0.66)	0.02 (0.02)	−1.19* (0.50)	0.01 (0.02)	−0.73* (0.36)	0.00 (0.01)
Early outcome score[e]	0.04 (0.03)	−0.00 (0.00)	0.57*** (0.03)	0.00*** (0.00)	0.12*** (0.02)	0.00 (0.00)	0.64*** (0.02)	−0.00 (0.00)
Middle childhood family variables[a]								
Average marital status	1.13 (0.92)	−0.07 (0.04)	0.39 (1.05)	−0.00 (0.02)	−1.06 (0.66)	−0.01 (0.02)	−0.60 (0.63)	0.02 (0.02)
Average maternal employment	0.06*** (0.02)	0.00 (0.00)	0.04* (0.02)	−0.00 (0.00)	−0.02 (0.02)	−0.00 (0.00)	−0.01 (0.01)	0.00 (0.00)
Average childhood income	0.09 (0.14)	−0.01 (0.01)	0.41* (0.16)	−0.00 (0.00)	−0.06 (0.12)	0.01 (0.00)	−0.27** (0.09)	−0.00 (0.00)
Average children in home	0.85* (0.35)	0.00 (0.01)	0.10 (0.41)	0.01 (0.01)	−0.53 (0.27)	0.00 (0.01)	−0.72** (0.23)	0.00 (0.01)
Classroom variables[a]								
Average teacher experience	0.14* (0.06)	0.00 (0.00)	−0.05 (0.05)	−0.00 (0.00)	−0.11** (0.04)	−0.00* (0.00)	0.04 (0.04)	−0.00 (0.00)
Average classroom quality	0.32** (0.10)	0.01 (0.01)	0.09 (0.12)	−0.00 (0.00)	−0.22** (0.07)	−0.00 (0.00)	−0.17* (0.08)	0.00 (0.01)

Note. $N = 1{,}364$. SSRS = Social Skills Rating System; CBCL = Child Behavior Checklist; PPVT = Peabody Picture Vocabulary Test; HOME = Home Observation for Measure of the Environment.

[a] Data from first, third, and fifth grades. [b] Data from 36 months of age. [c] Data from 6 to 54 months of age. [d] Data from 15 to 54 months of age. [e] Data from 54 months of age.

*$p < .05$. **$p < .01$. ***$p < .001$.

deviation above average displayed social skills levels that were 0.43 and 0.21 *SD* higher based on teacher and mother reports, respectively. Higher levels of teacher-reported parent involvement were also associated with fewer teacher-reported behaviors problems in children. Specifically, children whose parents scored a standard deviation above average displayed problem behaviors that were lower by 0.36 *SD* higher based on teacher reports. In sum, between-individual differences in average parent involvement across elementary school predicted higher social functioning and lower teacher-reported behavior problems on average from first through fifth grades. Finally, average levels of teacher-reported parent involvement did not predict growth in social skills and problem behaviors over time.

As in the maternal report models, several child, family and school characteristics were also linked to between-child differences in social skills and problem behaviors. Earlier ratings of social skills and problem behaviors from the 54-month time point were associated with average level scores in their corresponding subsequent ratings except for teacher-reported SSRS ratings. The number of minors in a household was positively linked with average levels of teacher-rated social skills and negatively related to maternal reports of behavior problems. Higher classroom quality was positively related to teacher-rated social skills and negatively related to both maternal- and teacher-reported behavior problems. More teacher experience was related to higher average teacher-rated social skills and lower average levels of teacher-rated problem behaviors.

Discussion

The purpose of this study was to consider the benefits of parent involvement for academic and social development in elementary school. In doing so, we aimed to address the current literature's reliance on cross-sectional designs and draw attention to these processes during middle childhood, which has been previously understudied. The results of the between-child analyses suggested that higher parent involvement, as reported by mothers and teachers, promotes better social skills, fewer problem behaviors, and is unrelated to average achievement across elementary school. However, between-child variation in the average levels of parent involvement was generally not predictive of growth in achievement and socioemotional functioning across elementary school. The exception was a finding that increases in average levels of teacher-reported parent involvement are related to less growth in reading scores. Our within-child analyses linked increases in maternal-rated parent involvement to improved teacher-rated social skills and declines in maternal- and teacher-rated problem behaviors. Increases in teacher-rated parent involvement were also related to enhancements in teacher-rated social skills and declines in problem behaviors. In contrast, within-child growth in achievement was not related to changes in either measure of parent involvement.

Achievement and Parent Involvement

Overall, the between- and within-child analyses of parent involvement and achievement revealed consistent findings. Greater engagement in a variety of parent involvement practices was largely unrelated to academic achievement, and improvements in parent involvement did not predict gains in any of the standardized achievement measures, except one possibly spurious finding between changes in teacher-reported parent involvement and decreased reading growth. Other past work has not detected significant associations between parent involvement and achievement (e.g., Keith et al., 1986; Okpala et al., 2001; Reynolds, 1992; White et al., 1992). Past findings of positive between-child associations of parent involvement and achievement may be artifacts of selection bias, whereby involved parents differed from less involved parents in a variety of ways such as in their motivation and beliefs about parenting, education, and their children's development.

Another potential explanation for this lack of association concerns the breadth of our measure of parent involvement. The current study used a measure containing widely accepted dimensions of parent involvement, such as the value parents place on education and the frequency with which they visit their child's school. However, Sheldon and Epstein (2005) suggested that parental involvement is most efficacious when it is subject specific, such that parents need to be involved within whichever domain they are targeting for improvement. Likewise, in past experimental work that linked significant increases in parent involvement with growth in literacy skills, the targeted parent involvement strategies focused specifically on promoting reading skills such as learning sight words and letter–sound recognition (DeCusati & Johnson, 2004). A recent meta-analysis found that parental expectations and aspirations were the factors most highly related to achievement (Fan & Chen, 2001). In the current study, parents were surveyed about their feelings and thoughts about education, but these items may have been too general or few in number to sensitively detect change in children's achievement. Another possibility is that parent involvement is more globally beneficial for children's academic performance in school (e.g., average grades, homework completion) but does not specifically promote achievement in any particular domain. Fan and Chen (2001) examined 25 studies of parent involvement and academic achievement and found that average correlations between parent involvement and children's grades or grade point average, both global indicators of school performance, tended to be moderate in size ($r = .33$), but average correlations between involvement and math and reading achievement scores were much lower ($r = .18$).

It should be noted that the only other study of within-child associations between parent involvement and student achievement, by Dearing et al. (2006), demonstrated a significant association between improvements in parent involvement across elementary school and improvements in literacy skills during elementary school. However, there are some notable differences in this study and ours. Dearing et al. had a much smaller ($N = 281$ vs. $N = 1{,}364$) and more socioeconomically homogenous sample, consisting only of low-income children enrolled in an intervention program. Children from low-income families face a higher likelihood of confronting multiple developmental risk factors across contexts than their more affluent counterparts (Friedman & Chase-Lansdale, 2002); thus, a positive factor, like parent involvement, may be especially protective. Moreover, Dearing and colleagues used parental reports that indicated whether or not they ever participated in certain school-based activities, whereas the current study included more detailed parent and teacher perspectives of parent involvement. Therefore, differences in the present study and the Dearing et al. study may be attributable to sample and measurement characteristics.

Additionally, the within-child negative finding in the current study among teacher reports of parent involvement and reading skills, although a weak association, is also supported by past work that has uncovered negative associations among facets of parent involvement and achievement outcomes (Milne et al., 1986; Sui-Chu & Willms, 1996). These researchers argued that parents may become increasingly involved in an effort to be proactive when their children experience academic difficulties. In the early years of elementary school, parents may tend to become more involved when their children are struggling in reading, but not math, because teachers focus substantially more on reading in first and third grades (Pianta, Belsky, Houts, Morrison, & the NICHD ECCRN, 2007). This negative association for reading growth, unlike math, may reflect teachers' attempts to elicit more parental involvement among parents of struggling readers during elementary school when classroom instruction heavily targets literacy skills.

Finally, the current study unexpectedly revealed small, negative associations between classroom quality and achievement in both the within- and between-child analyses. Pianta et al. (2008) also uncovered mixed findings between emotional climate and achievement, with the measure of emotional quality showing negative associations with literacy and math in first grade and more positive associations as elementary school progressed. Quality, which was characterized as positive emotional climate in the current study, was found to be consistently predictive of fewer problem behaviors and higher social skills between children.

Social Development and Parent Involvement

Contrary to the achievement findings, significant associations were detected repeatedly among parent involvement, social skills, and behavior problems in both between- and within-child models. In fact, increases in parents' involvement over time were related to concomitant increases in children's social skills and declines in problem behaviors for both reporters of involvement. Not only do the current findings extend past cross-sectional associations between parent involvement and socioemotional functioning, but these improvements in socioemotional skills may hold longer term implications for future academic performance and behavior (Bub, McCartney, & Willett, 2007; Li-Grining, Votruba-Drzal, & Maldonado-Carreño, under review). Additionally, these associations in the within-child models suggest that the relation between parent involvement and social skills is not merely the product of selection bias. Instead, these findings may reflect a growing consistency in social expectations across settings as children progress through school. As parents become more involved with their children's education, they may be more likely to communicate with school personnel about their child's adjustment and behavior in class. In turn, parents may become more informed about their children's social difficulties from teachers and school staff, and then subsequently address and reinforce more positive behaviors at home. Although communication among parents and teachers would be expected to benefit academic development as well, it is possible that teachers and parents discuss social and behavioral issues more frequently than academic concerns.

Limitations and Future Work

Although the present study offers many advantages with its longitudinal design and multimethod measurements, several limitations should be noted. Future studies of parent involvement could include more assessments or experimental designs to address issues of causality. The current analytic techniques limit the ability to make causal inferences, even when rigorous statistical techniques are employed. Moreover, early childhood interventions often include parent involvement as a key program component, such as the Infant Health and Development Program (McCormick, McCarton, Brooks-Gunn, Belt, & Gross, 1998), Head Start (Administration for Children and Families., 2007), and the School Development Program (Haynes, Comer, & Hamilton-Lee, 1988). However, with so many program features, it is difficult to distinguish the unique contribution of parent involvement to these gains. Future studies should more directly focus on parent involvement efforts to reduce potential confounding variables from additional experimental or intervention factors. Additionally, future parent involvement researchers should try to collect extensive and comprehensive data on a variety of the practices considered to be parent involvement. Instruments that survey a wide range of parent involvement practices may mask the parent behaviors that support growth in children's academic and social competence.

Results from the current study suggest that both maternal- and teacher-reported parent involvement have predictive value for children's social development. Although the patterns in associations across reporters were remarkably similar, there were several instances in which either parent or teacher reports predicted an outcome that the other report did not. Such findings lend support to the practice of triangulation in collecting parent involvement data.

Finally, findings from the current study are limited in their generalizability. The NICHD ECCRN and Duncan (2003) asserted that although attrition and nonrandom selection in the NICHD SECCYD sample inhibit the application of findings to broader populations, the sample also embodies unique strengths. Such strengths include the breadth of economic, ethnic, and geographic diversity embedded in the NICHD SECCYD sample.

Conclusion

Overall, the present findings suggest that parents continue to wield considerable influence on children's development as children progress through school. It is important for future work to explore parent behaviors that support children's achievement. In addition, further exploration of how parents and teachers may be jointly responding to children's social and behavioral skills could help to elucidate the potential benefits of parent involvement for social development. Investigation of the possible selection factors that motivate parent involvement would also be useful to inform policies and interventions.

References

Achenbach, T. M. (1991a). *Manual for the Child Behavior Checklist/4-18 and Profile*. Burlington: Department of Psychiatry, University of Vermont.

Achenbach, T. M. (1991b). *Manual for the Teacher Report Form and Profile*. Burlington: Department of Psychiatry, University of Vermont.

Acock, A. C. (2005). Working with missing values. *Journal of Marriage and Family*, *67*, 1012–1028.

Administration for Children and Families. (2007). Office of Head Start. Retrieved January 29, 2008, from http://www.acf.hhs.gov/programs/hsb/

Allison, P. D. (1990). Change scores as dependent variables in regression analysis. In C. C. Clogg (Ed.), *Sociological methodology* (pp. 93–114). Oxford, UK: Basil Blackwell.

Allison, P. D. (2008, September 8–9). *Missing data*. A presentation for Statistical Horizons workshop series, Chicago.

Berger, E. H. (1991). Parent involvement: Yesterday and today. *The Elementary School Journal*, *91*, 209–219.

Bronfenbrenner, U., & Ceci, S. J. (1994). Nature-nurture reconceptualized in developmental perspective: A Bioecological Model. *Psychological Review*, *101*, 568–586.

Bub, K. L., McCartney, K., & Willett, J. B. (2007). Behavior problem trajectories and first-grade cognitive ability and achievement skills: A latent growth curve analysis. *Journal of Educational Psychology*, *99*, 653–670.

Caldwell, B. M., & Bradley, R. H. (1984). The HOME inventory and family demographics. *Developmental Psychology*, *20*, 315–320.

Catsambis, S. (2001). Expanding knowledge of parental involvement in children's secondary education: Connections with high school seniors' academic success. *Social Psychology and Education*, *5*, 149–177.

Dearing, E., Kreider, H., Simpkins, S., & Weiss, H. B. (2006). Family involvement in school and low-income children's literacy: Longitudinal associations between and within families. *Journal of Educational Psychology*, *98*(4), 653–664.

Dearing, E., McCartney, K., Weiss, H. B., Kreider, H., & Simpkins, S. (2004). The promotive effects of family education for low-income children's literacy. *Journal of School Psychology*, *42*, 445–460.

DeCusati, C. L. P., & Johnson, J. E. (2004). Parents as classroom volunteers and kindergarten students' emergent reading skills. *Journal of Educational Research*, *97*, 235–246.

Duch, H. (2005). Redefining parent involvement in Head Start: A two-generation approach. *Early Child Development and Care*, *175*, 23–35.

Duncan, G. J., & Gibson-Davis, C. M. (2006). Connecting childcare quality to child outcomes: Drawing policy lessons from nonexperimental data. *Evaluation Review*, *30*, 611–630.

Duncan, G. J., Magnuson, K. A., & Ludwig, J. (2004). The endogeneity problem in developmental studies. *Research in Human Development*, *1&2*, 59–80.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Test–Revised Manual for Forms L and M*. Circle Pines, MN: American Guidance Service.

Englund, M. M., Luckner, A. E., Whaley, G. J. L., & Egeland, B. (2004). Children's achievement in early elementary school: Longitudinal effects of parental involvement, expectations, and quality of assistance. *Journal of Educational Psychology*, *96*, 723–730.

Fan, X., & Chen, M. (2001). Parental involvement and students' academic achievement: A meta-analysis. *Educational Psychology Review*, *13*, 1–22.

Fantuzzo, J., McWayne, C., & Perry, M. A. (2004). Multiple dimensions of family involvement and their relations to behavioral and learning competencies for urban, low-income children. *School Psychology Review*, *33*, 467–480.

Friedman, R. J., & Chase-Lansdale, P. L. (2002). Chronic adversities. In M. Rutter & E. Taylor (Eds), *Child and Adolescent Psychiatry* (4th ed., pp. 261–276). Oxford, UK: Blackwell.

Gonzalez-DeHass, A. R., Willems, P. P., & Holbein, M. F. D. (2005). Examining the relationship between parental involvement and student motivation. *Educational Psychology Review*, *17*, 99–123.

Gresham, F. M., & Elliott, S. N. (1990). *The social skills rating system*. Circle Pines, MN: American Guidance Service.

Haynes, N. M., Comer, J. P., & Hamilton-Lee, M. (1988). The school development program: A model for school improvement. *Journal of Negro Education*, *57*, 11–21.

Herrold, K., & O'Donnell, K. O. (2008). Parent and family involvement in education, 2006–07 school year, from the National Household Education Surveys Program of 2007. Institute of Education Sciences: National Center for Education Statistics. Retrieved September 21, 2008, from http://nces.ed.gov/pubs2008/2008050.pdf

Hill, N. E., Castellino, D. R., Lansford, J. E., Nowlin, P., Dodge, K. A., Bates, J. E., et al. (2004). Parent academic involvement as related to school behavior, achievement, and aspirations: Demographic variations across adolescence. *Child Development*, *75*, 1491–1509.

Izzo, C. V., Weissberg, R. P., Kasprow, W. J., & Fendrich, M. (1999). A longitudinal assessment of teacher perceptions of parent involvement in children's education and school performance. *American Journal of Community Psychology*, *27*, 817–839.

Keith, T. Z., Reimers, T. M., Fehrmann, P. G., Pottebaum, S. M., & Aubey, L. W. (1986). Parental involvement, homework, and TV times: Direct and indirect effects on high school achievement. *Journal of Educational Psychology*, *78*, 373–380.

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*, 1198–1202.

Machen, S. M., Wilson, J. D., & Notar, C. E. (2004). Parental involvement in the classroom. *Journal of Instructional Psychology*, *32*, 13–16.

McCormick, M. C., McCarton, C., Brooks-Gunn, J., Belt, P., & Gross, R. T. (1998). The Infant Health and Development Program: Interim summary. *Journal of Developmental & Behavioral Pediatrics*, *19*, 359–370.

McWayne, C., Hampton, V., Fantuzzo, J., Cohen, H. L., & Sekino, Y. (2004). A multivariate examination of parent involvement and the social and academic competencies of urban kindergarten children. *Psychology in the Schools*, *41*, 363–377.

Miedel, W. T., & Reynolds, A. J. (1999). Parent involvement in early intervention for disadvantaged children: Does it matter? *Journal of School Psychology*, *37*, 379–402.

Miller-Johnson, S., Maumary-Gremaud, A., & the Conduct Disorders Research Group. (1995). *Parent–Teacher Involvement Questionnaire: Teacher version*. Durham, NC: Duke University.

Milne, A. M., Myers, D. E., Rosenthal, A. S., & Ginsburg, A. (1986). Single parents, working mothers, and the educational achievement of school children. *Sociology of Education*, *59*, 125–139.

National Center for Educational Statistics. (1994). *School and Staffing Survey*. Washington, DC: U.S. Department of Education, Institute of Educational Sciences.

National Institute of Child Health and Human Development Early Child Care Research Network. (1993). *The NICHD Study of Early Child Care: A comprehensive longitudinal study of young children's lives* (ERIC Document Reproduction Service No. ED 353 087).

National Institute of Child Health and Human Development Early Child Care Research Network. (2002). The relation of global first grade classroom environment to structural classroom features and teacher and student behaviors. *The Elementary School Journal*, *102*, 367–387.

National Institute of Child Health and Human Development Early Child Care Research Network. (2003). Social functioning in first grade: Associations with earlier home and child care predictors and with current classroom experiences. *Child Development, 74*, 1639–1662.

National Institute of Child Health and Human Development Early Child Care Research Network. (2004). Does class size in first grade relate to children's academic and social performance or observed classroom processes? *Developmental Psychology, 40*, 651–664.

National Institute of Child Health and Human Development Early Child Care Research Network & Duncan, G. J. (2003). Modeling the impact of child care quality on children's preschool cognitive development. *Child Development, 74*, 1454–1475.

Okpala, C. O., Okpala, A. O., & Smith, F. E. (2001). Parental involvement, instructional expenditures, family socioeconomic attributes and student achievement. *Journal of Educational Research, 95*, 110–115.

Pianta, R., Belsky, J., Houts, R., Morrison, F., & the NICHD ECCRN. (2007). Opportunities to learn in America's elementary classrooms. *Science, 315*(5820), 1795–1796.

Pianta, R., Belsky, J., Vandergrift, N., Houts, R., Morrison, F., & the NICHD ECCRN. (2008). Classroom effects on children's achievement trajectories in elementary school. *American Educational Research Journal, 45*, 365–367.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Raudenbush, S. W., Bryk, A., Cheong, Y. F., Congdon, R., & du Toit, M. (2004). *HLM 6: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.

Reynolds, A. J. (1989). A structural model of first-grade outcomes for an urban, low socioeconomic status, minority population. *Journal of Educational Psychology, 81*, 594–603.

Reynolds, A. J. (1992). Comparing measures of parental involvement and their effects if a academic achievement. *Early Childhood Research Quarterly, 7*, 441–462.

Rimm-Kaufman, S. E., Pianta, R. C., Cox, M. J., & Bradley, R. H. (2003). Teacher-rated family involvement and children's social and academic outcomes in kindergarten. *Early Education and Development, 14*, 179–198.

Royston, P. (2005). Multiple imputation of missing values: Update of ice. *Stata Journal, 5*, 527–536.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147–177.

Sheldon, S. B., & Epstein, J. L. (2005). Involvement counts: Family and community partner-ships and mathematics achievement. *Journal of Educational Research, 98*, 196–206.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford, UK: Oxford University Press.

Sui-Chu, E. H., & Willms, J. D. (1996). Effects of parental involvement on eighth-grade achievement. *Sociology of Education, 69*, 126–141.

Supplee, L. H., Shaw, D., Hailstones, K., & Hartman, K. (2004). Family and child influences on early academic and emotion regulatory behaviors. *Journal of School Psychology, 42*, 221–242.

von Hippel, P. T. (2007). Regression with missing Ys: An improved strategy for analyzing multiple imputed data. *Sociological Methodology, 37*, 83–117.

White, K. R., Taylor, M. J., & Moss, V. D. (1992). Does research support claims about the benefits of involving parent in early intervention programs? *Review of Educational Research, 62*, 91–125.

Widaman, K. F. (2006). Missing data: What to do with or without them. *Monographs of the Society for Research in Child Development, 71*(3), 42–64.

Woodcock, R. W., & Johnson, M. B. (1989, 1990). *Woodcock–Johnson Psycho-Educational Battery–Revised*. Itasca, IL: Riverside.

Child Development, May/June 2010, Volume 81, Number 3, Pages 1006–1020

The Dynamics of Free Recall and Their Relation to Rehearsal Between 8 and 10 Years of Age

Martin Lehmann and Marcus Hasselhorn
German Institute for International Educational Research, Center for Education and Development

The present study longitudinally examined changes in recall in children between the ages of 8 and 10 years. Given the increasingly sophisticated use of memory strategies during this developmental period, correspondences between study and recall dynamics were of particular interest. Seventy-six children performed free-recall tasks on 5 occasions over a 2-year period. Video and audio analyses revealed that children tended to recall items successively from nearby serial positions. This so-called lag-recency effect was particularly pronounced when items from nearby serial positions were conjunctly rehearsed during study. Implications for understanding study–recall correspondences are discussed in relation to other developmental changes of this period including memory capacity.

Over the last four decades, knowledge of children's acquisition of memory strategies has dramatically increased. Today, research on memory strategies represents one of the most important areas in the field of cognitive development and the mechanisms involved in the emergence of deliberate strategy use are assumed to be a key issue in understanding human cognitive growth during childhood. From the very beginning, empirical approaches to memory strategies have focused on both sides of strategy usage: the learning behavior or input itself and recall consequences or output. Although the relation between input and output behavior seems to be critically related to developmental changes in memory proficiency (cf. Schneider & Pressley, 1997), this relation is not yet fully understood. It is, however, not until individuals recall the memorized material, that memory strategies reveal their quality and peculiarity (Schneider & Pressley, 1997; Waters & Andreassen, 1983). In the present investigation, data from an earlier study on children's applied memory-strategy behavior during study and subsequent free recall (Lehmann & Hasselhorn, 2007) were reanalyzed and observed learning behavior was related to a measure of output order developed by Kahana (1996) in his research on adults' free recall. This enabled a direct comparison of the developing processes of rehearsal behavior and recall behavior. The main goal of our analyses was to provide a more precise description of age-related recall dynamics and their relation to children's strategic behavior than has been previously presented.

Regarding the development of rehearsal, the seminal work of Ornstein and Naus (1985; Naus & Ornstein, 1983) focused on the direct consequences of different rehearsal strategies for recall and indicated that the spontaneous use of active rehearsal behavior should emerge between the ages of 8 and 10 years following a period of passive rehearsal usage (Ornstein, Baker-Ward, & Naus, 1988; Ornstein & Naus, 1978). Passive rehearsal strategies are characterized by repeating the to-be-learned items either once (labeling) or several times (rote repetition or single-word rehearsal) but not in combination with other items. In contrast, it is the very repetition of several combined items which characterizes active rehearsal (cumulative rehearsal). With respect to the recall consequences of these strategies, passive rehearsal has been found to lead to poor recall, in particular for items from the prerecency portion of a list, that is, items from the beginning and middle of the list (Hagen & Kingsley, 1968; Hagen, Meacham, & Mesibov, 1970). In contrast, active rehearsal has been found to lead to a general increase in recall performance as well as to account for the primacy effect and hence a pronounced recall of early list items (e.g., Ornstein &

This research was supported by a grant awarded by the German Research Foundation (Ha1452/4-1 and 4-3) to the second author.

Correspondence concerning this article should be addressed to Martin Lehmann, German Institute for International Educational Research, Center for Education and Development, Schlossstr. 29, D-60486 Frankfurt a. M., Germany. Electronic mail may be sent to lehmann@dipf.de.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8103-0024

Naus, 1978). The characteristics of serial position curves (i.e., recall probabilities as a function of list position) have thus always been used as an indicator of qualitative differences in children's spontaneous strategy use.

Further demonstrations of the recall consequences of different rehearsal strategies come from studies that directly control rehearsal behavior by instruction or by predetermined rehearsal patterns. Naus and Ornstein (1983), for example, used a "reading aloud" procedure in which participants rehearsed preset item combinations by reading them aloud. This procedure allowed for the control of rehearsal activity. Again, more active rehearsal behavior and therefore a joint rehearsal of different items resulted in increased recall.

A recent longitudinal study in our own lab (Lehmann & Hasselhorn, 2007) revealed that even young children profit from using cumulative rehearsal. In this study, children between the ages of 8 and 10 years performed three trials of a free-recall task on five occasions over a period of 2 years. When working on the free-recall task, participants were instructed to "learn aloud," that is, to verbalize whatever they were doing and thinking while memorizing the items. The authors examined each interstimulus interval and were in this way able to show that children used up to four different strategies during the list-learning process. Three of these strategies comprised rehearsal strategies such as labeling, single-word rehearsal, and cumulative rehearsal, while the fourth consisted of a simple form of association and elaboration. Developmental analyses revealed a gradual shift from labeling to cumulative rehearsal with increasing age. In addition, this fine-grained analysis allowed the identification of intralist strategy changes and revealed that cumulative rehearsal was frequently used at early list positions, whereas labeling dominated for later list portions. With regard to recall, cumulative rehearsal was found to be positively correlated with recall performance from the second measurement point (beginning of third grade) onward. In contrast, the remaining rehearsal strategies were, if at all, negatively related with recall performance.

These results coincide with two suggestions made by Cuvo (1975) several decades ago. First, he suggested that large rehearsal-set sizes (mean number of different items conjunctly rehearsed between two item presentations) would allow more items to be studied jointly and therefore to be strongly associated. Second, he suggested that the repeated reentering of previously presented items should result in both a greater permanence of these items in the short-term store and the creation of various associations between different items. Both suggestions have been confirmed to occur in older (sixth graders) but not younger children (second graders; e.g., Ornstein, Naus, & Stone, 1977). With respect to the behavior shown by the children assessed by Lehmann and Hasselhorn (2007), when the children were younger, they were not able to reenter items presented earlier in the list into rehearsal sets at later list portions, a behavior that was displayed when they were older. Similarly, Rundus (1971) demonstrated that adult participants tend to more frequently repeat items from the beginning of a list throughout the learning phase than items from later list positions, a behavior that again results in a notable primacy effect. A similar finding has been described by Tan and Ward (2000; Ward, 2002; Ward & Tan, 2004) who demonstrated that the noticeably better recall of items from the beginning of a list was due to the fact that these items were reentered into later rehearsal sets.

The approaches presented so far primarily focused on the recall consequences of specific learning behavior. While the examination of cumulative rehearsal behavior during study has generally been seen as an active process resulting in both a combination of different items within rehearsal sets and a transfer of earlier list items to rehearsal sets toward the end of the learning phase, corresponding recall analyses have received less attention. Information on recall sequences is, however, particularly important for identifying potential, underlying recall processes and, in turn, clarifying the relation between study and recall (Laming, 2006). With respect to the development of increasingly proficient strategies during input, one would expect a corresponding developmental change during output that may lead to more proficient recall behavior.

Bousfield's (1953) pioneering work on subjective organization during study was one of the first approaches to explicitly allow inferences about processes that mediate input and output behavior. In this approach, recall processes during output (e.g., subjective organization of the to-be-learned material) are assumed to reflect processes during input. Bousfield's (1953; Bousfield & Bousfield, 1966) methods paved the way for an enrichment of basic theory on memory strategies and learning. However, in contrast to the vast scope of knowledge on the relation between study and recall in organization, almost nothing is known about the processes directly linking study and recall in rehearsal. This is in part due to the lack of measures covering both the direct link between input and output and a process-oriented observation of output in free recall.

This issue has, in recent years, been addressed by Howard and Kahana (1999; Kahana, 1996) who showed in young adults that the temporal proximity of items during input influences the recall order of these items during output by studying the temporal item-presentation sequence at the time of encoding and the respective item output sequence at the time of recall. They found a strong tendency for items presented at neighboring positions during the learning phase to be successively recalled during the recall phase. In contrast, no such tendency was observed for items presented at more remote positions during the learning phase. They referred to this phenomenon as the *lag-recency effect*. This lag-recency effect differs from the end-of-list recency effect. While the end-of-list recency effect seems to be based on the high activation of items presented at or "transported" to the end of a list and therefore on the temporal proximity of study and recall (Howard & Kahana, 1999; Tan & Ward, 2000), the lag-recency effect refers to the proximity or strength of association between two successively presented items. During recall, the first of these two items to be recalled can operate as a retrieval cue for the second item. The observed recall transitions thus seem to be strongly influenced by temporally defined interitem associations (Kahana, Howard, & Polyn, 2008). This influence is reflected in both recall order and interresponse times (IRTs) and respectively quantified as the conditional response probability (CRP) and the conditional response latency (CRL; Kahana, 1996). As a function of lag, CRP describes the probability of items being recalled in succession given the distance (or lag) between them during the presentation phase. As indicated earlier, lag-CRP curves show a characteristic shape: Around the starting point, which is defined by the item to have just been recalled, items from nearby list positions (lags +1 and −1) show high probabilities of subsequently being recalled. With increasing lag (i.e., increasing distance between items on the list, e.g., lags −2, −3, . . . , and +2, +3, . . .), the probability of these items subsequently being recalled substantially decreases. Based on a reanalysis of data from large free-recall experiments (Murdock, 1962; Murdock & Metcalfe, 1978; Murdock & Okada, 1970), Kahana (1996) detected three basic characteristics of lag-CRP curves: (a) After recalling a specific item, the subsequently recalled item has a high probability of coming from a nearby list position; (b) the conditional response probabilities reveal an asymmetry, with the recall of items from previous list positions being less probable than the recall of items from subsequent list positions; and (c) the probability of recalls from nearby list positions is high at the beginning of the recall phase and subsequently decreases.

On the basis of these findings, Howard and Kahana (2002) proposed a *temporal-context model* that was able to explain contiguity effects in different free-recall tasks (e.g., Howard & Kahana, 1999; Kahana & Howard, 2005). In this model, recalling an item partially reinstates the context in which the item was presented and studied. This partial reinstatement of context serves as a retrieval cue for those items that shared a similar context during study. Since items from nearby list positions are most likely to have shared a similar context (partial), retrieval of this context results in close temporal recall of these neighboring items or, in other words, in the observed contiguity effect.

As mentioned earlier, little is known about the development of the relation between study and recall dynamics in children when implementing rehearsal activities. The present study thus had two main purposes. First, we aimed to increase understanding of general developmental changes in recall dynamics within a free-recall task and, second, to particularly explore the developing relation between strategic behavior and subsequent recall dynamics. To this end, we reanalyzed data from an earlier study by Lehmann and Hasselhorn (2007) to determine whether there are developmental changes in recall dynamics and whether these recall dynamics are due to children's rehearsal behavior.

With regard to the first purpose of examining developmental changes in recall dynamics, Lehmann and Hasselhorn (2007) and Ornstein and Naus (1985; Naus & Ornstein, 1983) have both shown developmental changes in rehearsal behavior and increases in recall performance. These findings, however, were again based on an examination of recall performance as a static end product of learning behavior rather than as a dynamic process in itself. Yet the aforementioned contiguity effect (e.g., Howard & Kahana, 1999) and the influence of semantic similarity on retrieval processes (e.g., Hasselhorn, 1992; Howard & Kahana, 2002) suggest that one has to analyze recall order when it comes to understanding recall as a subject-driven process. Such an analysis allows us to identify the consequences of preceding learning behavior and developmental processes in the functioning of episodic memory. Since the ability to encode ordinal structures and sequences of single events emerges very early in childhood, we suppose that associative effects, in terms of contiguity, can be observed within children's recall behavior.

Concerning the development of recall dynamics, the processes underlying study and recall can be inferred not only from the order in which items are recalled but also from the IRTs between them (Bousfield & Sedgewick, 1944; Pollio, 1974; Pollio, Richards, & Lucas, 1969). As soon as successive item retrievals are no longer independent of one another—thus fulfilling a main assumption of the temporal-context model (Howard & Kahana, 2002)—their IRTs are also assumed to be shorter as compared to the retrieval of almost unrelated items. In other words, we suppose that the stronger the association between two items or the greater the overlap between two contexts, the shorter the interval between the output of these items.

Concerning the second main purpose of more directly exploring the relation between strategic behavior and recall dynamics, we analyzed whether the peculiarities of children's rehearsal behavior identified in the study by Lehmann and Hasselhorn (2007) had differential effects on the corresponding recall dynamics. In one of the few previous studies on rehearsal-recall correspondences, Ornstein, Naus, and Stone (1977) analyzed the relation between the joint rehearsal of several words (i.e., cumulative rehearsal) and increased recall of items from early list positions. The authors applied Pellegrino's (1971) measure of intertrial repetition for subjective organization with units of size 2 and examined whether differences in input strategies (namely, rehearsal) were associated with differences in output organization. No differences in output organization were found for either different rehearsal strategies or for different age groups. However, results of two other studies by Ornstein, Naus, and Miller (1977) and Naus, Ornstein, and Aivano (1977) suggest that the joint rehearsal of several different items becomes increasingly important when a list lacks inherent (categorical-grouping) structure. Ornstein and Naus (1978, p. 95) suppose that under these conditions, "active rehearsal may serve an organizational function." Since Lehmann and Hasselhorn used semantically unrelated items and identified large qualitative and quantitative differences between strategies during study, we assumed that different forms of learning behavior would lead to considerable differences in subsequent recall behavior. In contrast to labeling and single-word rehearsal, cumulative rehearsal was explicitly assumed to represent a memory strategy that enables different items to share a very similar learning context. This shared context should, according to the temporal-context model (Howard & Kahana, 2002), result in an increased probability of close temporal recall for items sharing this context. In addition, this effect should be even more pronounced when items frequently share the same context in terms of repeatedly being selectively rehearsed.

Finally, Lehmann and Hasselhorn (2007) found that memory capacity constituted a strong factor underpinning both qualitative and quantitative developmental changes in active rehearsal behavior (see also Guttentag, 1984, 1989). Since memory capacity is assumed to be responsible for the retrieval of previously presented items that are subsequently available for rehearsal, we hypothesized that memory capacity should propel the association of items with one another during study and recall.

Method

Participants

Participants were 76 children (48 girls and 28 boys) who were assessed on five occasions over the course of a 2-year longitudinal study. The children attended various elementary schools in the area of Göttingen in Germany. They voluntarily participated (with parental permission) and were primarily from Caucasian middle-income families. At the beginning of the study, children were currently completing the latter half of second grade (*M* age = 8.3 years, *SD* = 6 months). Following their first assessment, children were tested on a further four occasions approximately 6 months apart. At the fifth measurement point they thus attended the latter half of fourth grade (*M* age = 10.2 years, *SD* = 6 months).

Materials and Procedure

Children studied three different lists of words and subsequently completed immediate free-recall tests. With the exception of the first measurement point, which lists comprised 14 items, the lists consisted of 12 highly concrete, familiar items from different semantic categories. Each list contained only one item from each category and the three lists were parallel with regard to the number of syllables. At each measurement point the three lists were administered in an identical and predetermined order. Individual stimulus items were simultaneously presented acoustically and visually at a rate of one item every 8 s. Acoustic presentation was performed using a CD player and visual presentation using 5 × 5 cm picture cards for each item, with the name of the object printed underneath its picture. Children were tested individually and both their learning and recall behavior were videotaped and digitally recorded.

At the beginning of each of the three trials, children were informed that they would be acoustically and visually presented with a number of items that were to be memorized and subsequently recalled in any order they wished. In addition, the children were instructed to think aloud while memorizing the items and to verbalize whatever they were thinking or doing (for a similar procedure, see Guttentag, Ornstein, & Siemens, 1987). List presentation was performed from a CD and by simultaneously placing the picture cards, one at a time, in front of the children. Each card was visible for an interval of 8 s until the subsequent item was acoustically presented by the CD and the card was replaced by the subsequent picture card. Following presentation of the entire list, children were instructed to recall as many items as possible. The recall phase ended after a 1-min interval in which the child was not able to recall any more words. In order to ensure that children were able to meet task requirements and to familiarize them with the thinking-aloud procedure, three practice trials (each including six items) were conducted at the first measurement point. At subsequent measurement points, only one practice trial was administered.

Lehmann and Hasselhorn (2007) assessed children's memory capacity with a digit span backward task, similar to the Wechsler Intelligence Scale for Children (Hamburg–Wechsler Intelligenztest fuer Kinder III [HAWIK–III]; Tewes, Rossmann, & Schallberger, 1999). In this task, digit sequences of increasing length were presented, starting with sequences of 2 one-syllable digits. At each sequence length, two series were presented acoustically at a rate of one digit every second. The digit span backward score was the maximal sequence length recalled in the correct backward order. If both sequences of the same length were accurately recalled, half a point was added to the score.

Coding and Scoring

For scoring and coding of strategic behavior, Lehmann and Hasselhorn (2007) used the audio and video recordings. During each interstimulus interval, children's behavior was coded according to five different strategy classifications: the three rehearsal strategies labeling (speaking aloud the name of one item only once), single-word rehearsal (repeating the name of one item several times), and cumulative rehearsal (speaking aloud a rehearsal set containing at least two items), as well as associative and elaborative learning behavior, and *no* observable learning behavior. At each measurement occasion, the percentages of occurrence of the different strategies were calculated. Two independent raters agreed on 89%, 88%, and 94% of strategy classification of labeling, single-word rehearsal, and cumulative rehearsal, respectively.

In addition children's verbalization of the to-be-learned items was assessed and coded and scored at each rehearsal opportunity. Two independent raters agreed on about 95% of the assessment of verbalizations. The few disagreements were resolved through inspection of the recordings and discussion.

For scoring and coding of recall behavior, again audio and video recordings were used. Two independent raters agreed on about 95% of assessment of number and order of items recalled and on about 90% of assessment of the time between each recall. Again, disagreements were resolved through inspection of the recordings and discussion.

Results

The previous analyses by Lehmann and Hasselhorn (2007) showed that cumulative rehearsal primarily occurred in early portions of the list. In contrast, the highest percentages for labeling were found in later list portions. With increasing age, however, children also showed incremental use of cumulative rehearsal in later list sections. Single-word rehearsal was used in a stable manner over both the course of the list-learning process and the five measurement points. Regarding the quality of rehearsal, mean rehearsal-set size, as measured by the number of different items included in an interstimulus interval, was observed to steadily increase over the five measurement points, with mean values of 1.11 (*SD* = 0.78), 1.38 (*SD* = 0.79), 1.64 (*SD* = 1.03), 1.90 (*SD* = 1.08), and 2.00 (*SD* = 1.15) for Measurement Points 1–5, respectively. In addition, when analyzing the data of only those children who actually engaged in cumulative rehearsal maximum rehearsal sets observed at the five measurement points were 3.8, 4.0, 4.7, 4.9, and 5.0 (range = 2–8) for Measurement Points 1–5, respectively.

At the five measurement points, children recalled a mean number of 5.25 (*SD* = 1.61), 5.55

(*SD* = 1.39), 5.85 (*SD* = 1.53), 6.28 (*SD* = 1.39), and 6.81 (*SD* = 1.43) items, respectively. Due to the fact that the three lists at the first measurement point contained two more items than the lists at later measurement points, recall percentages were analyzed. Recall percentages reliably increased with age over the five measurement points (for detailed information, see Lehmann & Hasselhorn, 2007).

Finally, Lehmann and Hasselhorn (2007) identified memory capacity to constitute a strong factor for the implementation of cumulative rehearsal. Memory capacity, assessed by the Digit Span Backward, increased reliably over the course of the elementary school years at intervals of 1 year (3.79, 3.88, 4.07, 4.24, and 4.43, for Measurement Points 1–5, respectively).

According to the goals of the present study, results are presented in two sections. In line with the first purpose of the study, the first section presents results on developmental changes in recall performance. In the second section, evidence of a correspondence between recall performance and strategic learning behavior are presented. Furthermore, evidence of a relation between selective mnemonic behaviors and the dynamics of free recall are reported. Finally, the influence of memory capacity on selective rehearsal behavior is demonstrated. Unless otherwise stated, all differences were significant at the $p < .05$ level.

Recall

Conditional response probability. So far, developmental changes in recall performance have almost exclusively been described using serial position curves. Serial position curves allow a valid inspection of recall performance with regard to the characteristics of the to-be-learned material and highlight developmental changes in recall performance. However, information on recall processes and the dynamics of recall remain greatly disregarded. In contrast, CRP curves allow for the identification of recall dynamics or more precisely of sequential contiguity. CRP is defined as "the probability of recalling item y after item x, conditionalized on recall of item y" (Kahana, 1996, p. 104). CRP is plotted as a function of lag, that is, as a function of the distance between two items within the list, reinstated at the time of recall. CRP as a function of lag (or lag-CRP) is calculated by dividing the number of transitions made to that lag (numerator) by the number of transitions that could have been made to that lag (denominator). The maximum lag is thus defined by the range of the list. In addition, the final position on the list does not allow for positive transitions and the first position does not allow for negative transitions.

To give an example: At the time of test, a person recalls six items out of a list of 10 items in the subsequent order (numbers correspond to the items' position within the list): 3, 4, 7, 5, 9, and 10. Given that there is no possible transition to the first item recalled, the following lags between recalled items can be identified: +1, +3, −2, +4, and +1. Considering the order of presentation of the list items, five transitions could have been made to lag +1 (to Items 4, 7, 5, 9, and 10), five transitions to lag +3 (to Items 4, 7, 5, 9, and 10), only four transitions to lag +4 (to Items 7, 5, 9, and 10) because there was no list item that could have been transitioned four list positions forward to List Item 4, and so on. Negative lags, that is, backward transitions at the time of testing, are subject to the same restrictions as forward transitions. In the given example, a negative transition to lag −2 could have been made three times (to List Items 4, 7, and 5). A negative transition to the lag −2, however, could not have been made to List Items 9 and 10 given that there were no items positioned after List Item 10.

We inspected the recall protocols from all five measurement points and analyzed developmental changes in the lag-recency effect. In the above-mentioned analyses performed by Kahana (1996), the first three recalled items were excluded in order to analyze free-recall data which had been freed from the recency effect. In the present study, preliminary analyses revealed that at Measurement Points 1–4, the probability of first recall for items from the last list position was at least twice as high as that of first recall for items from any other list position. Under the assumption that the recency effect in the present study was restricted to the first item recalled, this item was therefore excluded from all analyses. In addition, analyses only included recalls up to the fifth output. This was due to recall performance at the first measurement point in which the mean number of items recalled was approximately equal to 5. Hence, by limiting analyses to the early recalled items, we aimed (a) to balance children's recall performances that might emerge from large interindividual recall differences and (b) to avoid confounding of recall processes. Even if the lag-recency effect is observable at later recall transitions, this effect has been shown to be considerably smaller than that found at the beginning of the recall phase (Howard & Kahana, 1999). Since we were interested in recall orders of within-list items, repetitions and intrusions were excluded from analyses.

Figure 1 shows the lag-CRP functions for each of the five measurement points. As can be seen, the quantitative levels of the functions changed with age while the basic shape of the five curves remained highly similar. After recalling an item, the children tended to recall an item from a nearby list position. In order to analyze age-related differences in the lag-recency effect, we adopted the procedure developed by Kahana, Howard, Zaromb, and Wingfield (2002) and differentiated within the five lag-CRP functions between adjacent transitions (lags of +1 and −1) and remote transitions (lags of +3, +4, +5, and −3, −4, −5). An analysis of variance (ANOVA) with measurement points (1–5) and lags (adjacent and remote) as within-subject factors revealed a significant main effect of measurement point, $F(4, 296) = 7.54$, $MSE = 0.003$, $p < .001$; a significant main effect of lag, $F(1, 74) = 55.67$, $MSE = 0.01$, $p < .001$; and a significant interaction, $F(4, 296) = 3.19$, $MSE = 0.006$. While the form of the lag-CRP functions did not undergo large age-related changes, lag-CRP functions showed incremental steepness as the children increased in age. This was especially the case for the adjacent +1 lags that showed the highest probabilities at all measurement points but increasing probability with age and a peak at Measurement Point 5. In contrast, CRP functions for more remote lags (forward and backward) remained at almost the same low level. Tukey's honestly significant difference (HSD) tests supported these observations and revealed reliably higher probabilities for adjacent than remote lags at all measurement points. More important from a developmental perspective, however, was the finding that adjacent lags were far more pronounced at Measurement Point 5 than at Measurement Points 1, 3, and 4.

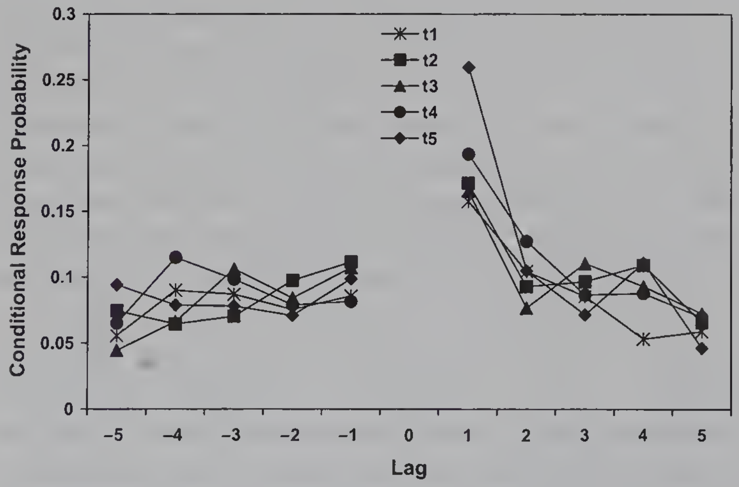


Figure 1. Conditional response probability curves as a function of lag when the children were 8.3 years (t_1), 8.9 years (t_2), 9.3 years (t_3), 9.9 years (t_4), and 10.2 years (t_5) old.

Conditional response latency. Associations between unrelated items can be formed when they share a similar episodic or rather temporally defined context. Associations are inferred by the tendency of one item evoking the other. Thus, associations between items are assumed to be responsible for recall sequences. These associations, however, not only appear to influence the order in which items are recalled but also the IRTs between them. In order to investigate this second characteristic of interitem associations, we recorded IRTs during children's recall and analyzed them conditional on recall transitions. Analyses of these so-called CRLs were comparable to those of the CRPs in that they included Output Positions 2–5 and followed the same requirements as the CRPs. IRT data were analyzed on the basis of computer-based visual display of the audio signals comprising vocalized recalls and pauses between recalls. IRTs were measured as the duration between the onset of a vocalized recall and the onset of a subsequent vocalized recall. Vocalization time was included in the IRTs given that vocalization of an item might itself overlap with children's search process for the next item. This assumption was confirmed by the tendency of some children to protract the endings of the items.

Figure 2 shows CRLs as a function of lag. An examination of these functions indicates that CRL's were shortest when successively recalled items were from nearby list positions. Quantitative analysis of the developmental latency data was again conducted by differentiating between adjacent transitions (lags +1 and −1) and remote transitions (lags +3, +4, +5, and −3, −4, −5). An ANOVA with measurement points (1–5) and lag (adjacent and remote) as within-subject factors only revealed a significant main effect of lag, $F(1, 30) = 6.71$, $MSE = 34{,}713{,}877$. Subsequent Tukey's HSD tests showed that latencies between items from adjacent input positions were shorter than those from remote input positions.

Relation Between Rehearsal Behavior and Recall Dynamics

In a reanalysis of rehearsal and recall data from a free-recall study by Murdock and Metcalfe (1978), Laming (2006, p. 1159) concluded that "recall and rehearsal result from a common underlying process." Since our CRP results reported earlier exclusively focused on the correspondence between the

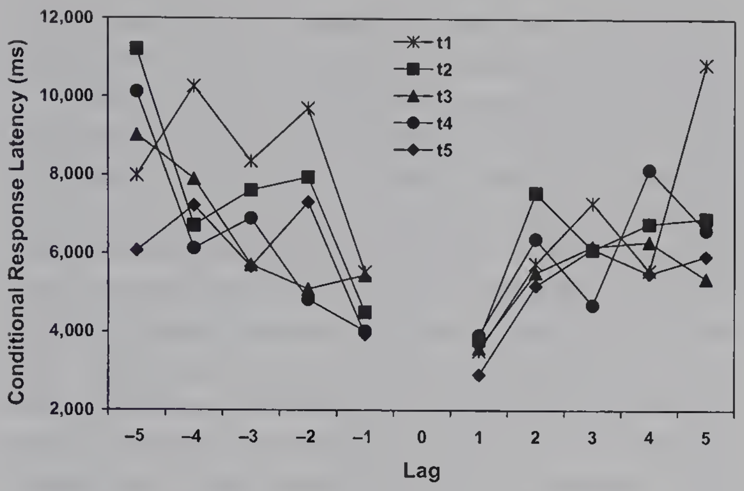


Figure 2. Conditional response latency curves as a function of lag when the children were 8.3 years (t_1), 8.9 years (t_2), 9.3 years (t_3), 9.9 years (t_4), and 10.2 years (t_5) old.

sequence of presentation and the sequence of recall, we aimed in a second step to examine whether and how learning behavior affects this correspondence. The children, who were given the possibility and even invited to actively memorize the lists of words, were observed by Lehmann and Hasselhorn (2007) to actually use three different rehearsal strategies (labeling, single-word rehearsal, and cumulative rehearsal). The following developmental analyses therefore focus on the influence of rehearsal behavior and in particular the impact of selective rehearsal of specific items or sets of items on the lag-recency effect. In the analysis of the lag-recency effect and thus also the recall order, we again differentiated between recalls from adjacent versus remote list positions.

Rehearsal behavior. Strategy development can be described in terms of both qualitative and quantitative changes (for detailed information, see Lehmann & Hasselhorn, 2007). Qualitative differences between rehearsal strategies are particularly characterized by the properties of the rehearsal sets involved, namely, rehearsal-set size. In the course of strategy development, children also show quantitative shifts in the frequency of the strategies which they apply. Specifically, children increasingly frequently and proportionally used superior strategies. In sum, advanced goal-directed strategic behavior (characterized by both qualitative and quantitative improvement) should also be reflected in resulting recall dynamics.

With respect to our data, correlation analyses revealed that mean rehearsal-set size was positively correlated with the mean probability of adjacent lags at Measurement Points 2, 3, and 4, $r(74) = .32$, $r(74) = .27$, and $r(74) = .41$, respectively. At Measurement Points 1 and 5, correlations were positive but nonsignificant, $rs(74) < .16$. Mean rehearsal-set sizes were not related to the average probability of remote lags at any of the five measurement points. Regarding quantitative aspects of rehearsal (i.e., the percentages of strategy use), the frequency of cumulative-rehearsal usage was positively related to the CRPs of adjacent lags at Measurement Points 2, 3, and 4, $r(74) = .28$, $r(74) = .28$, and $r(74) = .32$, respectively. At Measurement Points 1 and 5, correlations were again positive but nonsignificant, $rs(74) < .17$. Regarding the more passive rehearsal activities, only single-word rehearsal was positively related to the CRPs of remote lags at Measurement Point 2, $r(74) = .27$, and negatively related to the CRPs of adjacent lags, $r(74) = -.36$. As expected, proficient cumulative-rehearsal behavior during study, as characterized by rehearsal sets containing multiple items in close contiguity and high frequency in the execution of this behavior, was positively related to the recall of neighboring items. Significant relations between other rehearsal behavior and recall were only observed at Measurement Point 2, suggesting a smaller impact of other rehearsal strategies on systematic recall behavior.

The influence of selective rehearsal behavior on CRP and CRL. Selective rehearsal behavior can be defined as the selective repetition of specific items or a specific set of items. In a first step, we examined developmental changes in the occurrence of selective rehearsal behavior. In order to identify such behavior, we examined children's rehearsal protocols and classified pairs of items that were repeatedly rehearsed together. Again, we differentiated between pairs of items from adjacent list positions (lags +1 and −1) and pairs of items from remote list positions (lags +3, +4, +5, and −3, −4, −5). The first two columns in Table 1, respectively, display the number of joint repetitions of two items from nearby and remote list positions. A 2 (lag: adjacent and remote) × 5 (measurement points) ANOVA with repeated measures on both factors revealed main effects of lag, $F(1, 75) = 89.64$, $MSE = 0.28$, $p < .001$, and measurement point, $F(4, 300) = 36.00$, $MSE = 0.23$, $p < .001$. In addition, a significant Lag × Measurement Point interaction was found, $F(4, 300) = 39.43$, $MSE = 0.05$, $p < .001$. For both sets of item pairs, the mean number of joint repetitions increased as children got older. For items from adjacent serial positions, however, the mean number of joint repetitions increased more steeply than the mean number of joint rehearsals for more remote item pairs. These changes in rehearsal characteristics reflect a clear development toward increasing selective rehearsal behavior, in

Table 1

Mean Values (and Standard Deviations) for Number of Joint Rehearsals of Items From Adjacent and Remote Positions (General) and for Number of Joint Rehearsals of Items From Adjacent and Remote Positions (Effectively Recalled) at the Five Measurement Points (t_1–t_5)

	Number of joint rehearsals			
	General		Effectively recalled	
	Adjacent	Remote	Adjacent	Remote
t_1	0.04 (0.10)	0.05 (0.14)	1.33 (3.64)	0.68 (2.33)
t_2	0.44 (0.59)	0.12 (0.20)	2.86 (5.69)	0.64 (1.54)
t_3	0.63 (0.76)	0.22 (0.31)	3.76 (6.26)	1.61 (3.30)
t_4	0.82 (0.85)	0.29 (0.38)	4.51 (6.69)	1.78 (3.45)
t_5	0.91 (0.89)	0.34 (0.42)	5.95 (8.54)	1.97 (3.76)

particular in the rehearsal of items from nearby list positions. Subsequent Tukey's HSD tests confirmed this developmental pattern and revealed that with the exception of the first measurement point joint rehearsals of pairs of adjacent list items were more frequent than joint rehearsals of pairs of remote list items. Furthermore, the number of joint rehearsals of adjacent items steadily increased up to the beginning of fourth grade but not thereafter. Age-related differences with respect to the rehearsal of pairs of remote list items were only reliable in intervals of 1 year.

In order to compare processes of learning and recall, we analyzed the relation between the number of selective rehearsals and the recall order of these item pairs by once again examining children's recall protocols and differentiating between item pairs from adjacent and those from remote list positions. From Measurement Point 2 onward, a positive relation was observed between the number of rehearsals of adjacent list items and adjacent recall of these item pairs, $rs(74) > .24$. In contrast, the mean number of joint rehearsals for items from remote list positions was not related to adjacent recall of these item pairs at any of the five measurement occasions.

In a second step, we examined whether differences in the number of joint rehearsal of items affected subsequent consecutive recall of these items. Columns 3 and 4 in Table 1 present the mean number of joint rehearsals for items from adjacent and remote serial positions that were effectively recalled in succession. To control for differences in recall outcomes between adjacent and remote items, the number of joint repetitions of each adjacent and remote item pair was respectively related to the total number of recalls made for adjacent or remote item pairs. A two-way ANOVA with repeated measures on the factors lag (adjacent and remote) and measurement points (1–5) and the dependent variable "mean number of joint rehearsals of item pairs recalled" revealed main effects of lag, $F(1, 75) = 21.91$, $MSE = 47.68$, $p < .001$, and measurement point, $F(4, 300) = 13.44$, $MSE = 15.20$, $p < .001$. More importantly, however, the Lag × Measurement Point interaction also proved significant, $F(4, 296) = 4.38$, $MSE = 12.47$. With increasing age, joint rehearsal of item pairs became more and more effective. This was especially the case for items from adjacent serial positions and particularly apparent during fourth grade. Subsequent Tukey's HSD tests showed that from Measurement Point 2 onward, the mean number of joint rehearsals of effectively recalled adjacent item pairs was higher than the mean number of recalled remote item pairs. In addition, the number of joint rehearsals of effectively recalled adjacent item pairs was higher at Measurement Point 4 than at Measurement Point 1 and at Measurement Point 5 than at Measurement Points 1 and 3.

Since Lehmann and Hasselhorn (2007) reported that the percentages of cumulative-rehearsal use were not uniformly distributed across the list-learning process but were instead especially high at early- and middle-list positions, we further examined the distribution of pairs of adjacent items that were effectively recalled in succession. Given that that the group of item pairs from remote list positions was too heterogeneous to obtain an interpretable data pattern, distributions were examined for adjacent items only. At least 50 percent of successively recalled items from adjacent list positions (50%, 60%, 59%, 65%, and 68% for the five measurement occasions, respectively) were item pairs from the early list portion (i.e., Items 1–2, 2–3, 3–4, 4–5, and 5–6). Percentages for joint rehearsals of effectively recalled items from adjacent list positions were even more pronounced, with item pairs from early list positions amounting to at least 75% of the overall joint rehearsals of adjacent item pairs (75%, 79%, 90%, 79%, and 75% for the five measurement points, respectively).

We also tested the influence of joint rehearsal on recall latencies. Although most correlations were not significant, the highest correlation was interestingly found at Measurement Point 5, with the mean number of joint rehearsals of adjacent item pairs being related to short latencies between the successive recollection of these items, $r(70) = -.26$. While this latter finding should be interpreted cautiously, it suggests an increasing proficiency and effectiveness of selective rehearsal behavior at the end of Grade 4.

Memory capacity influence on selective rehearsal behavior. In the present study, additional analyses revealed positive correlations between memory capacity as measured by digit span backward and the number of joint rehearsals of items from neighboring serial positions at Measurement Points 2–5, *r*s(74) = .29, .37, .33, and .39, respectively. Correlation analyses between memory capacity and number of joint rehearsals of neighboring items that were effectively recalled in succession additionally revealed a positive relation at Measurement Points 2, 3, and 5, *r*s(74) = .33, .38, and .38, respectively. Findings for items from remote serial positions were, however, different. High memory capacity was, on the one hand, also positively correlated with the number of joint rehearsals of items from remote serial positions at Measurement Points 2–5, *r*s(74) = .36, .35, .30, and .42, respectively. On the other hand, no relation was found between memory capacity and the number of joint rehearsals of remote items that were effectively recalled in succession.

Discussion

While previous research (Lehmann & Hasselhorn, 2007; Ornstein & Naus, 1978) has highlighted the developmental changes in children's free-recall performance (age-related increase in active rehearsal leading to better recall of the to-be-learned material), the detailed method used in the present investigation yielded more insightful descriptions of recall dynamics in children and their developmental progression than those previously available. Recall dynamics are of particular interest because they offer an opportunity to relate sequential processes during study with processes during recall. In this concluding section, we thus successively consider developmental changes of sequential recall processes and the corresponding learning processes.

Sequential Effects During Recall

In the present analyses, we longitudinally examined age-related changes in recall dynamics following a list-learning process in children. Past analyses of rehearsal-recall observations have primarily focused on serial position curves that disregard sequential output positions. In contrast, the present analyses were based on a measure relating input order (i.e., the sequence of presentation during study) to output order (i.e., the sequence of recall). This detailed examination allowed us to analyze sequential dependencies across output positions and therefore revealed a much more process-oriented picture than has been yielded by previous analyses. We examined the so-called CRP that was introduced by Kahana (1996) and that gives the probability of recalling item *i* + lag after recalling item *i*. Analyses of the CRP revealed a high tendency for subsequently recalled items to come from nearby serial positions (lag-recency effect) and showed that this tendency is considerably higher for forward than for backward recalls.

In line with Kahana's (1996) findings, CRP analyses in the present study also revealed the lag-recency effect. Compared to the CRP curves generated for adults in studies by Kahana and his colleagues (Howard & Kahana, 1999; Kahana et al., 2002); however, we found a more pronounced asymmetry in the lag-recency effect in our sample of children aged between 8 and 10 years, with only small values and little variation in CRPs for backward recalls but large CRP values for forward recalls. This particularly applied to items from neighboring serial positions and much less so to items from remote serial positions. Furthermore, the present findings suggest that the lag-recency effect develops across the observed 2-year period: The effect was considerably more pronounced at the end of Grade 4 than at previous measurement points. In addition, recalls from more distant serial positions resulted in longer recall latencies and those from close serial positions in short recall latencies. In conclusion, the present findings suggest that, within a free-recall task comprising semantically unrelated material, interitem associations result from contextual or temporal proximity of the items. These interitem associations seem to result in associated items being recalled in very fast succession. The recall phenomena observed in the present study in 8- to 10-year-old children match recall phenomena in sort-recall tasks in children of the same age. Hasselhorn (1992), for example, analyzed 8- and 10-year-olds' interitem latencies in a sort-recall task and found that children's within-category latencies were shorter than their between-category latencies. Hence, recall of items sharing a similar temporal context (i.e., from adjacent serial positions) might be equated with the recall of items sharing a semantic context (i.e., from the same category) and, in turn, the recall of items from remote serial positions with the recall of items from different categories.

Relation Between Rehearsal Behavior and Recall Dynamics

Development in children's strategy use is characterized by their discovery of qualitatively novel strategies and concepts and by quantitative shifts in the frequency and efficiency with which they execute strategies (Siegler, 2006). The remarkable work of Ornstein and colleagues (e.g., Ornstein & Naus, 1978) has repeatedly shown that both the acquisition of advanced rehearsal strategies and qualitative changes in executing cumulative rehearsal (rehearsal-set size) are major contributors to recall performance and that these qualitative advancements increase with age. Recent research has further shown that quantitative improvement, that is, increasing frequency in the execution of cumulative rehearsal, also predicts recall performance (see Lehmann & Hasselhorn, 2007).

Nonetheless, previous analyses of rehearsal, have rarely addressed the issue of systematic change in recall dynamics as a consequence of developmental changes in rehearsal behavior. In allowing an assessment of sequential dependencies in retrieval, the present analyses permitted to demonstrate that both qualitative and quantitative developmental changes in rehearsal were related to sequential characteristics of recall. Specifically, increasing quality in rehearsal behavior (rehearsal-set size) and increasing frequency (percentages) of cumulative-rehearsal usage at Measurement Points 2–4 were related to the probability of recalling items from nearby serial list positions. These results suggest that, when implemented as an active goal-directed behavior, cumulative rehearsal leads to adjacent items being related during study and subsequently recalled in succession. These results differ from those reported by Ornstein, Naus, and Stone (1977) in a rehearsal-training study with second and sixth graders. Here, Pellegrino's (1971) measure of intertrial repetition for subjective units of size 2 did not yield a significant effect of rehearsal strategy on output organization for either of the two age groups. In contrast to the present study, however, Ornstein, Naus, and Stone (1977) used the same list of words with different orders across several trials. If indeed rehearsal is characterized by a process that constructs retrievable units from neighboring items, then reordering items in each trial might result in order interferences between the different trials and thus obscure subjective organization or subjective clustering.

The observed cessation of the described effects after the fourth measurement point suggests that qualitative and quantitative changes in children's rehearsal behavior beyond this developmental stage potentially might not be detected by such global measures as rehearsal-set size or the number of implemented strategies. This assumption is supported by the obvious limitations of global measures of rehearsal and in particular of rehearsal-set size. Naus et al. (1977), for example, found a clear limitation of rehearsal-set size, with sets larger than approximately three to four items failing to additionally enhance recall. Hence, even if children are able to compile large rehearsal sets, they might find it difficult to do so throughout the entire list-learning process. If children therefore find it difficult to continually and conjunctly rehearse a large number of items and additionally have the impression that they do not especially profit from large rehearsal sets during recall, then other mechanisms beyond solely compiling large rehearsal sets should come into play.

Selective Rehearsal and Its Influences on Recall Dynamics

According to the assumption that a fine-grained approach is needed to detect developmental changes in the relation between specific rehearsal behavior and specific recall behavior, we analyzed the joint rehearsal of particular items. Our aim was therefore to identify whether these sequential dependencies were due to selective and focused learning behavior with respect to these items. By analyzing the composition of rehearsal sets, we found that when children grew older (in the course of Grades 3 and 4), they showed conjunct rehearsal of items from nearby serial positions far more often than conjunct rehearsal of items from remote serial positions.

In focusing on pairs of items that were rehearsed together and subsequently recalled in succession, the present study contributed a further method for validating rehearsal-recall matches. The examination of pairs of jointly rehearsed items and their recall correspondences offers the possibility to identify study-recall patterns that include both qualitative and quantitative aspects of rehearsal behavior: The more often pairs of items are rehearsed together in different interstimulus intervals, the higher the percentage of cumulative-rehearsal usage according to its definition (cumulative rehearsal is defined as an average rehearsal set of at least two different items; e.g., Guttentag et al., 1987). In addition, the more often items (here, item pairs) are reentered into cumulative-rehearsal sets for additional study, the larger the rehearsal sets and the more frequent and various the opportunities for interitem associations become (see also Cuvo, 1975).

Examining the number of conjoint reentries in interstimulus intervals of pairs of items that were additionally effectively recalled thus allowed us to establish that successively recalled items were far more frequently studied together when they came from nearby as opposed to remote serial positions. This effect became even more pronounced with increasing age.

Implications for the Development of Rehearsal–Recall Correspondences

The current approach to examining developmental changes in study-recall behavior lead us to suggest the following study-recall processes in children when working on a free-recall task: The experimentally predetermined item presentation sequence can be seen as a kick-off for associating items presented in close proximity in terms of the temporal-context model proposed by Howard and Kahana (2002). When children are allowed to engage in strategic behavior, neighboring items are rehearsed together. This mechanism of repetition increases the interitem associations between these items as a result of which the items can be kept available for further rehearsal and more easily retrieved. This characteristic is particularly noticeable in younger children who show less proficient forms of cumulative rehearsal. In a training study by Naus et al. (1977), third graders were instructed to rehearse the currently presented item with the two previously presented items. While children were able to implement cumulative-rehearsal behavior, most of them showed a rather inflexible rehearsal pattern and rehearsed each currently presented item with the items from the first two serial positions. With increasing age, however, children's rehearsal behavior becomes more flexible. At these later stages, both a large rehearsal-set size and a high proportion of cumulative-rehearsal usage presumably lead to a continuous reentering of previously presented items into rehearsal sets and consequently, to an increase in interitem associations and an intermixture of new and old items generating various retrieval cues for later recall. This is in accordance with an observation by Ornstein, Naus, and Liberty (1975), some decades ago, who found in eighth-grade children's rehearsal protocols many previously presented items as being rehearsed together in several later rehearsal sets.

Developmental increase in memory capacity would appear to be an essential factor establishing this developmental change. Memory capacity is assumed to be responsible for the retrieval of previously presented items that are subsequently available for rehearsal and association with one another. In the present study, additional analyses revealed that memory capacity exerted its influence on both rehearsal of item pairs from adjacent and remote serial positions during study but not on rehearsal of item pairs from remote serial positions during recall. Specifically, high memory capacity was positively correlated with the ability to compile many items (from different serial positions) into rehearsal sets and to maintain the process of joint rehearsal of (associated) items over a prolonged period. Regarding the link between study and recall, however, the observed effect of memory capacity on study was only mirrored in effects on recall for items from neighboring positions. Items that are successively presented and rehearsed may therefore be strongly associated with one another and in turn be successively recalled. Recalling a specific item seems to reinstate the learning context, which serves as a retrieval cue for other items with a similar context (Kahana et al., 2008). The present study reveals that this study-recall pattern develops during childhood. More particularly, in the present sample, this correspondence was found to increase up to the age of approximately 10 years. Presumably, the high processing demands of multi-item rehearsal constrain in younger children the ability to rehearse many different items conjointly, that is, in an active, goal-directed strategic behavior that should prompt construction of retrievable units of adjacent items. However, if memory demands are reduced during study (e.g., by keeping the items visible during study) even younger children (Grade 3) are able to use cumulative rehearsal in an efficient manner and to rehearse different items together (see Guttentag et al., 1987; Ornstein, Medlin, Stone, & Naus, 1985).

Limitations and Future Research

While the results of the present study suggest a clear lag-recency effect in children that increases with age, they should be interpreted with some caution. On the one hand, the lag-recency effect is usually tested using a larger number of lists. Using three lists, as was the case in the present study, causes results to be less reliable. On the other hand, previous studies have shown the lag-recency effect to be rather robust and have demonstrated that it is still to be found in adults at late output positions (Howard & Kahana, 1999). Due to this robustness and given that the CRPs found in the present study closely resemble those reported in previous studies with adults as well as showing the same properties as described in these studies, using CRPs to assess recall processes in children appears to represent an adequate option.

The present analyses indicate that when children work on a free-recall task, processes taking place at the time of study and at the time of test seem to be compatible. However, this correspondence seems to be mediated by children's memory capacity, which itself increases with age. Employing complex memory strategies such as cumulative rehearsal requires enormous mental effort. The proficient use of cumulative rehearsal (i.e., high-quality rehearsal over a prolonged period of time) thus seems to necessitate high memory capacity. It is well established that older children have larger memory spans and are therefore able to store more information. As shown by Guttentag et al. (1987) and Lehmann and Hasselhorn (2007), smaller memory demands (e.g., by strong experience with the task) or higher memory capacity allow children to more extensively make use of cumulative rehearsal or to use it at all. However, we know very little about the sources of these capacity differences and their specific impact on rehearsal. In light of Cowan's (1997) review of several causes of developmental changes in children's memory capacity, including, for example, more storage space, faster processing speed, improved attention, and less memory loss over time, its impact on rehearsal would seem to be far from straightforward. In order to more precisely understand the factors underlying the impact of memory capacity on rehearsal processes, more detailed analyses both during study and recall are necessary. Future research should extend the present analyses by specifying the locus of capacity advantages in older children within the rehearsal process. It may, for example, be argued that differences in rehearsal rate explain why some children use and profit from an active rehearsal process while others do not (e.g., Schneider & Bjorklund, 1998). However, the assumption that the amount of time it takes to rehearse or recite the to-be-learned material represents the defining factor in explaining capacity differences does not comply with findings that emphasize the role of chunk limits in storage capacity differences (e.g., Cowan, 2001). In the present study, the identification of strongly associated neighboring items would seem to favor the latter explanation. In sum, it may prove fruitful to examine whether the assessment of different aspects of capacity reveals different aspects of rehearsal and subsequent rehearsal-recall correspondences.

Finally, the present results raise the question concerning the existence of further, qualitatively different developmental steps toward a sophisticated mastery of free-recall tasks. Addressing this question may include a search for additional internally driven sources that account for study-recall correspondences beyond memory capacity. It might, for example, be assumed that under the premise of capacity influences, children would be able to allocate their resources to control processes or behavior that additionally structures the to-be-learned material both during study and recall. The developmental changes in memory-related activities may, however, additionally be driven by external factors as experience and schooling practice. For instance, Coffman, Ornstein, McCall, and Curran (2008) showed that first-grade teachers' provision of metacognitive information and request for deliberate remembering during instruction had a critical impact on their pupils' use of mnemonic techniques and ability to generalize training in sorting and clustering strategies. Thus, from our point of view, memory capacity should be regarded as a necessary but not sufficient source for rehearsal development.

References

Bousfield, W. A. (1953). The occurence of clustering in the recall of randomly arranged associates. *Journal of Genetic Psychology*, *49*, 229–240.

Bousfield, A. K., & Bousfield, W. A. (1966). Measurement of clustering and of sequential constanties in repeated free recall. *Psychological Reports*, *19*, 935–942.

Bousfield, W. A., & Sedgewick, C. H. (1944). An analysis of sequences of restricted associated responses. *Journal of General Psychology*, *30*, 149–165.

Coffman, J. L., Ornstein, P. A., McCall, L. E., & Curran, P. J. (2008). Linking teachers' memory-relevant language and the development of children's memory skills. *Developmental Psychology*, *44*, 1640–1654.

Cowan, N. (1997). The development of working memory. In N. Cowan (Ed.), *The development of memory in childhood* (pp. 163–199). East Sussex, UK: Psychology Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–185.

Cuvo, A. J. (1975). Developmental differences in rehearsal and free recall. *Journal of Experimental Child Psychology*, *19*, 265–278.

Guttentag, R. E. (1984). The mental effort requirement of cumulative rehearsal: A developmental study. *Journal of Experimental Child Psychology*, *37*, 92–106.

Guttentag, R. E. (1989). Age differences in dual-task performance: Procedures, assumptions, and results. *Developmental Review, 9*, 146–170.

Guttentag, R. E., Ornstein, P. A., & Siemens, L. (1987). Children's spontaneous rehearsal: Transitions in strategy acquisition. *Cognitive Development, 2*, 307–326.

Hagen, J. W., & Kingsley, P. R. (1968). Labeling effects in short-term memory. *Child Development, 39*, 113–121.

Hagen, J. W., Meacham, J. A., & Mesibov, G. (1970). Verbal learning, rehearsal, and short-term memory. *Cognitive Psychology, 1*, 47–58.

Hasselhorn, M. (1992). Task dependency and the role of category typicality and metamemory in the development of an organizational strategy. *Child Development, 63*, 202–214.

Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 923–941.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology, 46*, 269–299.

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory and Cognition, 24*, 103–109.

Kahana, M. J., & Howard, M. W. (2005). Spacing and lag effects in free recall of pure lists. *Psychonomic Bulletin & Review, 12*, 159–164.

Kahana, M. J., Howard, M. W., & Polyn, S. M. (2008). Associative retrieval processes in episodic memory. In H. L. Roediger (Ed.), *Learning and memory: A comprehensive reference* (pp. 476–490). Oxford, UK: Academic Press.

Kahana, M. J., Howard, M. W., Zaromb, F., & Wingfield, A. (2002). Age dissociates recency and lag-recency effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 530–540.

Laming, D. (2006). Predicting free recalls. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 1146–1163.

Lehmann, M., & Hasselhorn, M. (2007). Variable memory strategy use in children's adaptive intratask learning behavior: Developmental changes and working memory influences in free recall. *Child Development, 78*, 1068–1082.

Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology, 64*, 482–488.

Murdock, B. B., & Metcalfe, J. A. (1978). Controlled rehearsal in single-trial free recall. *Journal of Verbal Learning & Verbal Behavior, 17*, 309–324.

Murdock, B. B., & Okada, R. (1970). Interresponse times in singletrial free recall. *Journal of Verbal Learning & Verbal Behavior, 86*, 263–267.

Naus, M. J., & Ornstein, P. A. (1983). The development of memory strategies: Analysis, questions, and issues. In M. T. H. Chi (Ed.), *Trends in memory development research: Vol. 9. Contributions to human development* (pp. 1–30). Basel, Switzerland: S. Karger.

Naus, M. J., Ornstein, P. A., & Aivano, S. (1977). Developmental changes in memory: The effects of processing time and rehearsal instructions. *Journal of Experimental Child Psychology, 23*, 237–251.

Ornstein, P. A., Baker-Ward, L., & Naus, M. J. (1988). The development of mnemonic skill. In F. E. Weinert & M. Perlmutter (Eds.), *Memory development: Universal changes and individual differences* (pp. 31–51). Hillsdale, NJ: Erlbaum.

Ornstein, P. A., Medlin, R. G., Stone, B. P., & Naus, M. J. (1985). Retrieving for rehearsal: An analysis of active rehearsal in children's memory. *Developmental Psychology, 21*, 633–641.

Ornstein, P. A., & Naus, M. J. (1978). Rehearsal Processes in children's memory. In P. A. Ornstein (Ed.), *Memory development in children* (pp. 69–99). Hillsdale, NJ: Erlbaum.

Ornstein, P. A., & Naus, M. J. (1985). Effects of the knowledge base on children's memory strategies. *Advances in Child Development and Behavior, 19*, 113–148.

Ornstein, P. A., Naus, M. J., & Liberty, C. (1975). Rehearsal and organizational processes in children's memory. *Child Development, 46*, 818–830.

Ornstein, P. A., Naus, M. J., & Miller, T. D. (1977). The effects of list organization and rehearsal activity on children's free recall. *Child Development, 48*, 292–295.

Ornstein, P. A., Naus, M. J., & Stone, B. P. (1977). Rehearsal training and development differences in memory. *Developmental Psychology, 13*, 15–24.

Pellegrino, J. W. (1971). A general measure of organization in free recall for variable unit size and internal sequential consistency. *Behavior Research Methods & Instrumentation, 3*, 241–246.

Pollio, H. R. (1974). *The psychology of symbolic activity*. Reading, MA: Addison-Wesley.

Pollio, H. R., Richards, S., & Lucas, R. (1969). Temporal properties of category recall. *Journal of Verbal Learning and Verbal Behavior, 8*, 529–536.

Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology, 89*, 63–77.

Schneider, W., & Bjorklund, D. F. (1998). Memory. In W. Damon, D. Kuhn, & R. S. Siegler (Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (pp. 467–521). New York: Wiley.

Schneider, W., & Pressley, M. (1997). *Memory development between 2 and 20* (2nd ed.). Mahwah, NJ: Erlbaum.

Siegler, R. S. (2006). Microgenetic analyses of learning. In W. Damon & R. M. Lerner (Series Eds.) & D. Kuhn & R. S. Siegler (Vol. Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (6th ed., pp. 464–510). Hoboken, NJ: Wiley.

Tan, L., & Ward, G. (2000). A recency-based account of the primacy effect in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1589–1625.

Tewes, U., Rossmann, K., & Schallberger, U. (1999). *Hamburg-Wechsler-Intelligenztest für Kinder III (HAWIK III)*. Göttingen, Germany: Hogrefe.

Ward, G. (2002). A recency-based account of the list length effect in free recall. *Memory and Cognition*, *30*, 885–892.

Ward, G., & Tan, L. (2004). The effect of the length of to-be-remembered lists and intervening lists on free recall: A reexamination using overt rehearsal. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 1196–1210.

Waters, H. S., & Andreassen, C. (1983). Children's use of memory strategies under instruction. In M. Pressley & J. R. Levin (Eds.), *Cognitive strategies: Developmental, educational, and treatment-related issues.* (pp. 3–24). New York: Springer.

Child Development, May/June 2010, Volume 81, Number 3, Pages 1021–1023

Manuscripts Accepted for Publication

Sex Differences in Salivary Cortisol, Alpha-Amylase, and Psychological Functioning Following Hurricane Katrina, Jacob Vigil, David Geary, Douglas Granger, and Mark Flinn.

Exposure to 9/11 Among Youth and Their Mothers in New York City: Enduring Associations With Mental Health and Sociopolitical Attitudes, Elizabeth Gershoff, J. Lawrence Aber, Angelica Ware, and Jennifer Kotler.

Exploring Posttraumatic Growth in Children Impacted by Hurricane Katrina: Correlates of the Phenomenon and Developmental Considerations, Ryan Kilmer and Virginia Gil-Rivas.

Sierra Leone's Former Child Soldiers: A Follow-Up Study of Psychosocial Adjustment and Community Reintegration, Theresa Betancourt, Ivelina Borisova, Timothy Phillips, Robert T. Brennan, Theodore Whitfield, Marie de la Soudiere, John Williamson, and Stephen Gilman.

Impact of Maternal Posttraumatic Stress Disorder and Depression Following Exposure to the September 11 Attacks on Preschool Children's Behavior, Claude M. Chemtob, Yoko Nomura, Khushmand Rajendran, Rachel Yehuda, Deena Schwartz, and Robert Abramovitz.

Tsunami, War, and Cumulative Risk in the Lives of Sri Lankan Schoolchildren, Claudia Catani, Abigail H. Gewirtz, Elizabeth Wieling, Elisabeth Schauer, Thomas Elbert, and Frank Neuner.

Growing Pains: The Impact of Disaster-Related and Daily Stressors on the Psychological and Psychosocial Functioning of Youth in Sri Lanka, Gaithri Fernando, Kenneth Miller, and Dale Berger.

Children With Disabilities in the Context of Disaster: A Social Vulnerability Perspective, Lori Peek and Laura M. Stough.

Typologies of Family Functioning and Children's Adjustment During the Early School Years, Melissa Sturge-Apple, Patrick Davies, and E. Mark Cummings.

Defining Best Practice in Care and Protection of Children in Crisis-Affected Settings: A Delphi Study, Alastair Ager, Lindsay Stark, Bree Akesson, and Neil Boothby.

Posttraumatic Resilience in Former Ugandan Child Soldiers, Fionna Klasen, Gabriele Oettingen, Judith Daniels, Manuela Post, Catrin Hoyer, and Hubertus Adam.

The Effects of the 1999 Turkish Earthquake on Young Children: Analyzing Traumatized Children's Completion of Short Stories, Elif Celebi Oncu and Aysegul Metindogan Wise.

The Kenyan Political Conflict and Children's Adjustment, Mumbe Kithakye, Amanda Morris, Andrew Terranova, and Sonya Myers.

An Eye for Detail: An Event-Related Potential Study of the Rapid Processing of Fearful Facial Expressions in Children, Petra Vlamings, Lisa M. Jonkman, and Chantal Kemner.

On the Progression and Stability of Adolescent Identity Formation, Wim Meeus, Rens van de Schoot, Loes Keijsers, Seth J. Schwartz, and Susan Branje.

Dogs, Bogs, Labs, and Lads: What Phonemic Generalizations Indicate About the Nature of Children's Early Word-Form Representations, Erik Thiessen and Meagan N. Yee.

Investigating the Development of Core Cognitive Skills in Autism: A 3-Year Prospective Study, Elizabeth Pellicano.

Home Media and Children's Achievement and Behavior, Sandra Hofferth.

The Effects of Maltreatment and Neuroendocrine Regulation on Memory Performance, Dante Cicchetti, Fred A. Rogosch, Mark L. Howe, and Sheree L. Toth.

Interracial and Intraracial Contact, School-Level Diversity, and Change in Racial Identity Status Among African American Adolescents, Tiffany Yip, Eleanor K. Seaton, and Robert M. Sellers.

Child Care and the Development of Behavior Problems Among Economically Disadvantaged Children in Middle Childhood, Elizabeth

Votruba-Drzal, Rebekah Coley, Carolina Maldonado Carreno, Christine Li-Grining, and P. Lindsay Chase-Lansdale.

An Ecological Approach to Promoting Early Adolescent Mental Health and Social Adaptation: Family-Centered Intervention in Public Middle Schools, Elizabeth Stormshak, Arin M. Connell, Marie-Helene Veronneau, Michael W. Myers, Thomas J. Dishion, Kathryn Kavanagh, and Allison S. Caruthers.

Assessing the Effects of a Work-Based Antipoverty Program for Parents on Youth's Future Orientation and Employment Experiences, Rachel Kaplan, Vonnie C. McLoyd, Kelly Purtell, and Aletha Huston.

Children's Classroom Engagement and School Readiness Gains in Pre-Kindergarten, Nina C. Chien, Carollee Howes, Margaret Burchinal, Robert Pianta, Sharon Ritchie, Donna Bryant, Richard Clifford, Diane Early, and Oscar Barbarin.

Early Elementary School Adjustment of Maltreated Children in Foster Care: The Roles of Inhibitory Control and Caregiver Involvement, Katherine Pears, Philip A. Fisher, Jacqueline Bruce, Hyoun K. Kim, and Karen Yoerger.

Children's Understanding of Ordinary and Extraordinary Minds, Jonathan Lane, Henry Wellman, and E. Margaret Evans.

What Paradox? Referential Cues Allow for Infant Use of Phonetic Detail in Word Learning, Christopher Fennell and Sandra R. Waxman.

Contagious Yawning in Autistic and Typical Development, Molly Helt, Inge-Marie Eigsti, Peter J. Snyder, and Deborah A. Fein.

Tolerance of Practices by Muslim Actors: An Integrative Social-Developmental Perspective, Maike Gieling, Jochem Thijs, and Maykel Verkuyten.

Structure and Strategies in Children's Educational Television: The Roles of Program Type and Learning Strategies in Children's Learning, Deborah L. Linebarger and Jessica Taylor Piotrowski.

Bridging the Gap: Solving Spatial Means–Ends Relations in a Locomotor Task, Sarah Berger, Karen E. Adolph, and Alisan E. Kavookjian.

Sex Categorization Among Preschool Children: Increasing Utilization of Sexually Dimorphic Cues, Kerri Johnson, Leah Lurye Murphy, and Louis G. Tassinary.

Gender Differences in Spatial Ability of Young Children: The Effects of Training and Processing Strategies, David Tzuriel and Gila Egozi.

Disasters, Victimization, and Children's Mental Health, Kathryn Becker-Blease, Heather Turner, and David Finkelhor.

Do Adolescents and Parents Reconstruct Memories About Their Conflict as a Function of Adolescent Attachment? Matthew Dykas, Susan S. Woodhouse, Katherine B. Ehrlich, and Jude Cassidy.

Cognitive Flexibility in Drawings of Bilingual Children, Esther Adi-Japha, Jennie Berberich-Artzi, and Afaf Libnawi.

The Contributions of Numerosity and Domain-General Abilities to School Readiness, Lynn Fuchs, David Geary, Donald Compton, Douglas Fuchs, Carol Hamlett, and Joan Bryant.

Prospective Associations Between Peer Victimization and Aggression, Jamie Ostrov.

Relations Between Physiological and Cognitive Regulatory Systems: Infant Sleep Regulation and Subsequent Executive Functioning, Annie Bernier, Stephanie M. Carlson, Stéphanie Bordeleau, and Julie Carrier.

Maternal Discussions of Mental States and Behaviors: Relations to Emotion Situation Knowledge in European American and Immigrant Chinese Children, Stacey Doan and Qi Wang.

Reading Minds: The Relation Between Children's Mental State Knowledge and Their Metaknowledge About Reading, Serena Lecce, Silvia Zocchi, Adriano Pagnin, Paola Palladino, and Mele Taumoepeau.

Do Black and Hispanic Children Benefit More From Preschool? Understanding Differences in Preschool Effects Across Racial Groups, Daphna Bassok.

Adolescents' Development of Skills for Agency in Youth Programs: Learning to Think Strategically, Reed Larson and Rachel Angus.

Preverbal Infants Anticipate That Food Will Be Brought to the Mouth: An Eye Tracking Study of Manual Feeding and Flying Spoons, Olga Kochukhova and Gustaf Gredebäck.

Race Salience and Essentialist Thinking in Racial Stereotype Development, Kristin Pauker, Nalini Ambady, and Evan Apfelbaum.

Differing Levels of Gender Salience in Preschool Classrooms: Effects on Children's Gender Attitudes and Intergroup Bias, Lacey J. Hilliard and Lynn S. Liben.

Comprehension of Infrequent Subject–Verb Agreement Forms: Evidence From French-Learning Children, Geraldine Legendre, Isabelle Barriere, Louise Goyet, and Thierry Nazzi.

Two-Year-Olds Use the Generic/Nongeneric Distinction to Guide Their Inferences About Novel Kinds, Susan Graham, Samantha Nayer, and Susan Gelman.

Goal Neglect and Working Memory Capacity in 4- to 6-Year-Old Children, Stuart Marcovitch, Janet J. Boseovski, Robin J. Knapp, and Michael J. Kane.

The Social Costs of Academic Success Across Ethnic Groups, Thomas E. Fuller-Rowell and Stacey N. Doan.

Children's Reasoning About Disclosing Adult Transgressions: Effects of Maltreatment, Child Age, and Adult Edentity, Thomas Lyon, Elizabeth A. Ahern, Lindsay A. Malloy, and Jodi A. Quas.

A Developmental Perspective on Executive Function, John R. Best and Patricia Miller.

Pathways to Mathematics: Longitudinal Predictors of Performance, Jo-Anne LeFevre, Lisa Fast, Sheri-Lynn Skwarchuk, Brenda Smith-Chant, Jeffrey Bisanz, Deepthi Kamawar, and Marcie Penner-Wilger.

Development in Children's Comprehension of Linguistic Register, Laura Wagner, Maia Greene-Havas, and Rebecca Gillespie.

How 15 Hundred Is Like 15 Cherries: Effect of Progressive Alignment on Representational Changes in Numerical Cognition, Clarissa Thompson and John Opfer.

Toddlers' Prosocial Behavior: From Instrumental to Empathic to Altruistic Helping, Margarita Svetlova, Sara R. Nichols, and Celia A. Brownell.

Movement Planning Reflects Skill Level and Age Changes in Toddlers, Yu-ping Chen, Rachel Keen, Kerstin Rosander, and Claes von Hofsten.

Young Children Selectively Avoid Helping People With Harmful Intentions, Amrisha Vaish, Malinda Carpenter, and Michael Tomasello.

Danish Majority Children's Reasoning About Exclusion Based on Gender and Ethnicity, Signe Juhl Moller and Harriet Tenenbaum.

How Does It Look? Level 2 Perspective Taking at 36 Months of Age, Henrike Moll and Andrew Meltzoff.

Emotion Discourse, Social Cognition, and Social Skills in Children With and Without Developmental Delays, Rachel M. Fenning, Bruce L. Baker, and Jaana Juvonen.

A Monozygotic Twin Differences Study of Friends' Aggression and Children's Adjustment Problems, Frank Vitaro, Mara Brendgen, Michel Boivin, Stephane Cantin, Ginette Dionne, Richard E. Tremblay, Alain Girard, and Daniel Perusse.

Children of Katrina: Lessons Learned About Post-Disaster Symptoms and Recover Partners, Mindy Kronenberg, Tonya Hansel, Adrianne Brennan, Beverly Lawrason, Howard Osofsky, and Joy Osofsky.

Teacher–Child Relationships and the Development of Academic and Behavioral Skills During Elementary School: A Within- and Between-Child Analysis, Carolina Maldonado-Carreño and Elizabeth Votruba-Drzal.

Object Similarity Bootstraps Young Children to Action-Based Verb Extensions, Etsuko Haryu, Mutsumi Imai, and Hiroyuki Okada.

Can Maltreated Children Inhibit True and False Memories for Emotional Information? Mark Howe, Sheree Toth, and Dante Cicchetti.

Child Development, May/June 2010, Volume 81, Number 3, Page 1024

Errata

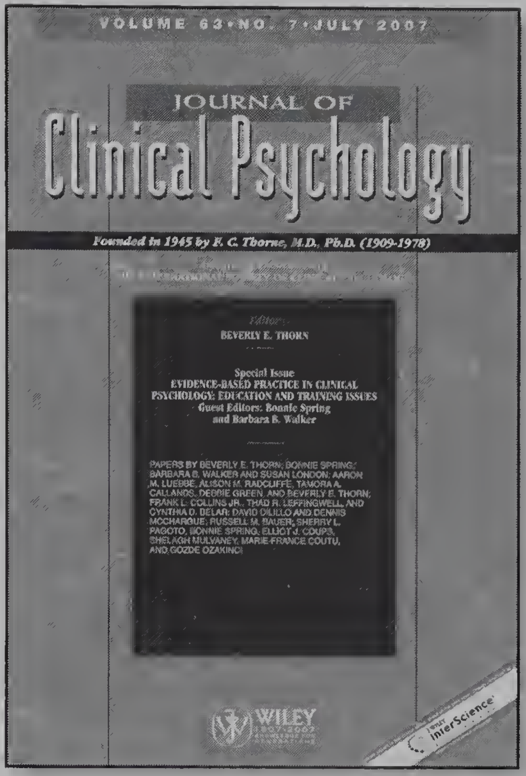
Steinberg, L., Graham, S., O'Brien, L., Woolard, J., Cauffman, E., & Banich, M. (2009). Age differences in future orientation and delay discounting. *Child Development*, *80*, 28–44.

On page 44 in the Appendix of the article, the first sentence of the scoring instructions should read: ''Reverse score items 1, 3, 4, 6, 8, 11, 13, and 15, so that higher scores indicate a stronger future orientation.''

Lagattuta, K. H., Nucci, L., & Bosacki, S. L. (2010). Bridging theory of mind and the personal domain: Children's reasoning about resistance to parental control. *Child Development*, *81*, 616–635.

This article contained an error in the presentation of Figures 6 and 7 on pages 628–629. Although the figure captions are correct, they do not correspond to the accompanying bar graphs. The bar graph in Figure 7 (not Figure 6) displays data for identity-related explanations, and the bar graph in Figure 6 (not Figure 7) presents data for moral-evaluative explanations.

© 2010, Copyright the Author(s)
Journal Compilation © 2010, Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2010/8103-0026



VOLUME 63 • NO. 7 • JULY 2007

JOURNAL OF
Clinical Psychology

Founded in 1945 by F. C. Thorne, M.D., Ph.D. (1909-1978)

Editor:
BEVERLY E. THORN

Special Issue
EVIDENCE-BASED PRACTICE IN CLINICAL
PSYCHOLOGY: EDUCATION AND TRAINING ISSUES
Guest Editors: Bonnie Spring
and Barbara B. Walker

PAPERS BY BEVERLY E. THORN; BONNIE SPRING; BARBARA B. WALKER AND SUSAN LONDON; AARON M. LUEBBE, ALISON M. RADCLIFFE, TAMORA A. CALLANDS, DEBBIE GREEN, AND BEVERLY E. THORN; FRANK L. COLLINS JR., THAD R. LEFFINGWELL, AND CYNTHIA D. BELAR; DAVID DILILLO AND DENNIS MCCHARGUE; RUSSELL M. BAUER; SHERRY L. PAGOTO, BONNIE SPRING, ELLIOT J. COUPS, SHELAGH MULVANEY, MARIE-FRANCE COUTU, AND GOZDE OZAKINCI

WILEY

InterScience

JOURNAL OF
Clinical Psychology

Edited by
BEVERLY E. THORN

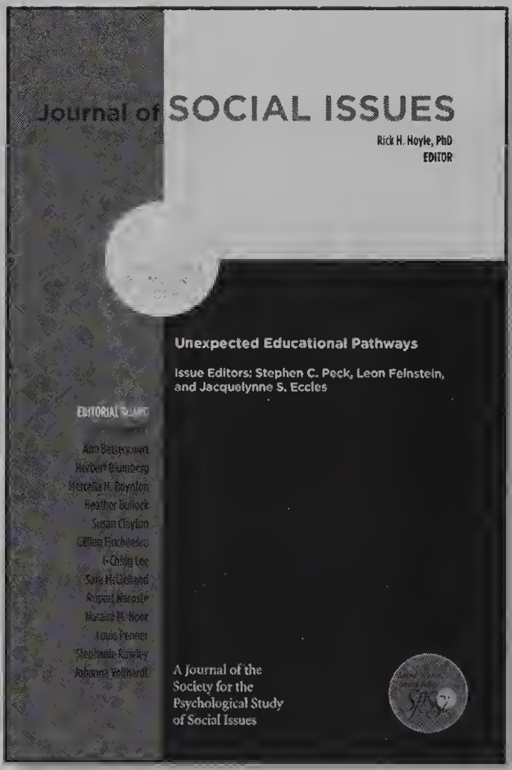
Founded in 1945, the ***Journal of Clinical Psychology*** is a peer-reviewed forum devoted to research, assessment, and practice. Published eight times a year, the ***Journal*** includes research studies; articles on contemporary professional issues, single case research; brief reports (including dissertations in brief); notes from the field; and news and notes. In addition to papers on psychopathology, psychodiagnostics, and the psychotherapeutic process, the journal welcomes articles focusing on psychotherapy effectiveness research, psychological assessment and treatment matching, clinical outcomes, clinical health psychology, and behavioral medicine.

From time to time, the ***Journal*** publishes Special Sections, featuring a selection of articles related to a single particularly timely or important theme; individuals interested in Guest Editing a Special Section are encouraged to contact the Editors.

WILEY-
BLACKWELL

For more information and to subscribe online visit
www.wileyinterscience.com/journal/jclp

558012



Journal of SOCIAL ISSUES
Rick H. Hoyle, PhD
EDITOR

Unexpected Educational Pathways
Issue Editors: Stephen C. Peck, Leon Feinstein, and Jacquelynne S. Eccles

A Journal of the Society for the Psychological Study of Social Issues

Journal of
SOCIAL
ISSUES

Edited by
SHERI R. LEVY

Published on behalf of The Society for the Psychological Study of Social Issues (SPSSI)

Journal of Social Issues (JSI) brings behavioral and social science theory, empirical evidence, and practice to bear on human and social problems. Each issue of the journal focuses on a single topic — recent issues, for example, have addressed poverty, housing, and health; privacy as a social and psychological concern; youth and violence; and the impact of social class on education.

A subscription to ***JSI*** also includes a full subscription to SPSSI's two other journals, ***Analyses of Social Issues and Public Policy (ASAP)*** and ***Social Issues and Police Review (SIPR)***.

WILEY-BLACKWELL

For more information and to subscribe online visit
www.interscience.wiley.com/journal/jsi

Notice to Contributors

S1U8

Child Development publishes empirical, theoretical, review, applied, and policy articles reporting research on child development. Published by the interdisciplinary Society for Research in Child Development (SRCD), the journal welcomes relevant submissions from all disciplines. Further information is available at http://www.srcd.org/cd.html.

Types of Articles

Child Development considers manuscripts in formats described below. Inquiries concerning alternative formats should be addressed to the Editor prior to submission. All submissions are expected to be no more than 40 manuscript pages, including tables, references, and figures (but excluding appendices). Authors should provide a justification if the submission is substantially longer. Unless the editor finds that justification compelling, the submission will be returned to the author for shortening prior to editorial review.

Empirical articles comprise the major portion of the journal. To be accepted, empirical articles must be judged as being high in scientific quality, contributing to the empirical base of child development, and having important theoretical, practical, or interdisciplinary implications. Reports of multiple studies, methods, or settings are encouraged, but single-study reports are also considered. Empirical articles will thus vary considerably in length, but should be no longer than 40 manuscript pages; text and graphics should be as concise as material permits. All modes of empirical research are welcome.

Empirical reports are reserved for short cutting-edge empirical papers that are no longer than 4,000 words in length (including text, tables, footnotes, appendices), which advance research and knowledge in an area through noteworthy findings and/or new methods. For manuscripts that require longer descriptions of methods and results, authors should use the Empirical article format.

Reviews focus on past empirical and/or on conceptual and theoretical work. They are expected to synthesize or evaluate a topic or issue relevant to child development, should appeal to a broad audience, and may be followed by a small number of solicited commentaries.

Essays describe original concepts, methods, trends, applications, and theories; these may also be accompanied by solicited commentaries.

Child Development and … are articles that provide readers with tutorials about some new concept or academic specialty pertinent to research in child development. These papers should review the major definitions, methods, and findings of the concept or specialty and discuss past or potential links to child development.

From another perspective is a format in which papers on a focal topic, written by different authors, are published simultaneously. Papers represent diverse perspectives (e.g., authors whose work represents different populations; different disciplines; different theories, methods, or analytic tools). In some cases, calls for submissions on particular topics will be disseminated through SRCD (via e-mail or SRCD publications), and submissions will undergo normal editorial review. In some cases, a submitted manuscript (e.g., an empirical article) may be selected as a lead article for this format, with invited commentaries providing additional perspectives. The editors also welcome suggestions from readers for topics for this format.

Manuscript Submission

Child Development invites for consideration manuscripts that are neither identical to nor substantially similar to work published or under review elsewhere. In the submission cover letter, please provide details about other published or submitted papers having substantial overlap (including data sets) with the new CD submission to enable editors to judge whether the new submission is sufficiently distinct from other work to warrant consideration. Please note if the paper is posted on a website, see http://www. srcd.org/web posting. html. Editors retain the right to reject manuscripts that do not meet established ethical standards for research or dissemination.

Manuscripts should be submitted electronically to the Child Development Online Submission Site at www.srcd.org /CDsubmit/ as a Word or WordPerfect file. Please also submit a cover letter that contains the name(s) of the author(s) and affiliation(s), and the street address, telephone, fax, and electronic mail address of the corresponding author. A corresponding author's submission to *Child Development* implies that all coauthors have agreed to the content and form of the manuscript and that the ethical standards of SRCD have been followed (see the *Child Development* website or pp. 283–284 of the 2000 SRCD Directory). Any financial interest or conflict of interest must be explained to the Editor in the cover letter. The corresponding author is responsible for informing all coauthors, in a timely manner, of manuscript submission, editorial decisions, reviews, and revisions. Please note that if your submission (or its revision) is accepted for publication, each author must submit a signed Exclusive License Form (http://www.srcd.org/Documents/Publications/CD/cdcopyright.pdf) before the manuscript can enter production. Planning ahead to meet this requirement is especially important for manuscripts with multiple authors.

The manuscript file should be formatted with double-spaced, 12-point type, and should include a single paragraph abstract of 100–120 words. Please follow all guidelines on format, style, and ethics provided in the Publication Manual (6th ed.) of the American Psychological Association. Figures included with initial submissions will not be returned. Therefore, please submit only electronic files or copies of figures. If color artwork is submitted, and the authors believe color art is necessary to the presentation of their work, the submissions letter should indicate that one or more authors or their institutions are prepared to pay the substantial costs associated with color art reproduction. Authors should keep a copy of all correspondence, files, and figures to guard against loss.

Manuscript Review

If you have any questions about your submission, please inquire at cdev@srcd.org or call (734) 926-0615. Each manuscript is handled by the Editor or an Associate Editor who consults with one or more Consulting Editors and/or ad hoc reviewers who have relevant expertise. To ensure blind review, cover sheets are removed before review; authors should avoid including any other information about identity or affiliation in submissions. Copies of the submission and associated correspondence are retained in the SRCD archives. For accepted manuscripts, authors are required to prepare a 300–500 layperson's summary for public dissemination purposes. Details are provided to authors as part of final processing.

There is no charge for publication in *Child Development* unless tabular or graphic materials exceed 10% of the total number of pages. Charges are also levied for changes in proofs other than correction of printer's errors. Any inquiries relating to charges or business matters (including reprint orders) should be addressed to Wiley-Blackwell, *Child Development*, Production Coordinator, 350 Main Street, Malden, MA 02148, (781) 388-8200.

Inquiries and suggestions regarding editorial policy may be addressed to: Dr. Jeffrey J. Lockman, Editor, *Child Development*, Department of Psychology, Tulane University, 2007 Percival Stern Hall, New Orleans, LA 70118-5698, lockman@tulane.edu. *Please do not send electronic or paper manuscripts to the above address.* Doing so will delay the processing of your manuscript.

Note to NIH Grantees

Pursuant to NIH mandate, Society through Wiley-Blackwell will post the accepted version of Contributions authored by NIH grantholders to PubMed Central upon acceptance. This accepted version will be made publicly available 12 months after publication. For further information, see www.wiley.com/go/nihmandate.

MONOGRAPHS OF THE SOCIETY FOR RESEARCH IN CHILD DEVELOPMENT

Recent

DEPRIVATION-SPECIFIC PSYCHOLOGICAL PATTERNS: EFFECTS OF INSTITUTIONAL DEPRIVATION
- *Michael Rutter, Edmund J. Sonuga-Barke, Celia Beckett, Jennifer Castle, Jana Kreppner, Robert Kumsta, Wolff Schlotz, Suzanne Stevens, and Christopher A. Bell* (Serial Na. 295, 2010).

A DYNAMIC CASCADE MODEL OF THE DEVELOPMENT OF SUBSTANCE-USE ONSET
- *Kenneth A. Dodge, Patrick S. Malone, Jennifer E. Lansford, Shari Miller, Gregory S. Pettit, and John E. Bates* (Serial Na. 294, 2009).

FLEXIBILITY IN EARLY VERB USE: EVIDENCE FROM A MULTIPLE-*N* DIARY STUDY
- *Letitia R. Naigles, Erika Hoff, and Donna Vear* (Serial Na. 293, 2009).

MARITAL CONFLICT AND CHILDREN'S EXTERNALIZING BEHAVIOR: INTERACTIONS BETWEEN PARASYMPATHETIC AND SYMPATHETIC NERVOUS SYSTEM ACTIVITY
- *Mona El-Sheikh et al.* (Serial Na. 292, 2009).

THE EFFECTS OF EARLY SOCIAL-EMOTIONAL AND RELATIONSHIP EXPERIENCE ON THE DEVELOPMENT OF YOUNG ORPHANAGE CHILDREN
- *The St. Petersburg–USA Orphanage Research Team* (Serial Na. 291, 2008).

UNDERSTANDING MOTHER-ADOLESCENT CONFLICT DISCUSSIONS: CONCURRENT AND ACROSS-TIME PREDICTION FROM YOUTHS' DISPOSITIONS AND PARENTING
- *Nancy Eisenberg et al.* (Serial Na. 290, 2008).

DEVELOPING OBJECT CONCEPTS IN INFANCY: AN ASSOCIATIVE LEARNING PERSPECTIVE
- *David H. Rakison and Gary Lupyan* (Serial Na. 289, 2008).

THE GENETIC AND ENVIRONMENTAL ORIGINS OF LEARNING ABILITIES AND DISABILITIES IN THE EARLY SCHOOL YEARS
- *Yulia Kovas, Claire M. A. Haworth, Philip S. Dale, and Robert Plomin* (Serial Na. 288, 2007).

THE PRESERVATION OF TWO INFANT TEMPERAMENTS INTO ADOLESCENCE
- *Jerome Kagan, Nancy Snidman, Vali Kahn, and Sara Towsley* (Serial No. 287, 2007).

CHILDREN'S QUESTIONS: A MECHANISM FOR COGNITIVE DEVELOPMENT
- *Michelle M. Chouinard* (Serial Na. 286, 2007).

BEST PRACTICES IN QUANTITATIVE METHODS FOR DEVELOPMENTALISTS
- *Kathleen McCartney, Margaret R. Burchinal, and Kristen L. Bub* (Serial No. 285, 2006).

FOUNDATIONS OF SELF-AWARENESS: AN EXPLORATION THROUGH AUTISM
- *R. Peter Hobson, Gayathri Chidambi, Anthony Lee, and Jessica Meyer* (Serial No. 284, 2006).

OUR CHILDREN TOO: A HISTORY OF THE BLACK CAUCUS
OF THE SOCIETY FOR RESEARCH IN CHILD DEVELOPMENT, 1973–1997
- *Diana T. Slaughter-Defoe, Aline M. Garrett, and Algea O. Harrison-Hale* (Serial Na. 283, 2006).

PARENTAL SUPPORT, PSYCHOLOGICAL CONTROL, AND BEHAVIORAL CONTROL:
ASSESSING RELEVANCE ACROSS TIME, CULTURE, AND METHOD
- *Brian K. Barber, Heidi E. Stolz, and Joseph A. Olsen* (Serial Na. 282, 2005).

BEING HURT AND HURTING OTHERS: CHILDREN'S NARRATIVE ACCOUNTS
AND MORAL JUDGMENTS OF THEIR OWN INTERPERSONAL CONFLICTS
- *Cecilia Wainryb, Beverly A. Brenl, and Sonia Matwin* (Serial Na. 281, 2005).

CHILDHOOD SEXUAL ASSAULT VICTIMS: LONG-TERM OUTCOMES AFTER TESTIFYING IN CRIMINAL COURT
- *Jodi A. Quas, Gail S. Goodman, Simona Ghetti, Kristen Weede Alexander, Robin Edelstein, Allison D. Redlich, Ingrid M. Cordon, and David P. H. Jones* (Serial Na. 280, 2005).

THE EMERGENCE OF SOCIAL COGNITION IN THREE YOUNG CHIMPANZEES
- *Michael Tomasello and Malinda Carpenter* (Serial Na. 279, 2005).

TRAJECTORIES OF PHYSICAL AGGRESSION FROM TODDLERHOOD TO MIDDLE CHILDHOOD:
PREDICTORS, CORRELATES, AND OUTCOMES
- *NICHD Early Child Care Research Network* (Serial No. 278, 2004).

CONSTRAINTS ON CONCEPTUAL DEVELOPMENT: A CASE STUDY OF THE ACQUISITION OF
FOLKBIOLOGICAL AND FOLKSOCIOLOGICAL KNOWLEDGE IN MADAGASCAR
- *Rita Astuti, Gregg E. A. Solomon, and Susan Carey* (Serial No. 277, 2004).

ORIGINS AND EARLY DEVELOPMENT OF HUMAN BODY KNOWLEDGE
- *Virginia Slaughter, Michelle Heron, Linda Jenkins, and Elizabeth Tilse* (Serial Na. 276, 2004).

MOTHER CHILD CONVERSATIONS ABOUT GENDER: UNDERSTANDING THE ACQUISITION OF ESSENTIALIST BELIEFS
- *Susan A. Gelman, Marianne G. Taylor, and Simone P. Nguyen* (Serial Na. 275, 2004).

WILEY-
BLACKWELL